Prepared for submission to JHEP

Subatomic Physics: the Notes[1]

C.P. Burgess

Department of Physics & Astronomy, McMaster University
and Perimeter Institute for Theoretical Physics

Contents

1 The story so far... 2
- 1.1 Prequel: substructure and atoms 2
- 1.2 Units and scales 19
- 1.3 Lies, Damn Lies, and Measurement Errors 23
- 1.4 Relativistic kinematics 30

2 Calculational tools I 39
- 2.1 Conserved quantities 39
- 2.2 Decays: general properties 41
- 2.3 Scattering: general properties 55

3 Calculational tools II 65
- 3.1 Classical two-body scattering 65
- 3.2 Quantum potential scattering 75
- 3.3 Perturbation theory and the Born approximation 92

4 Nucleon substructure 100
- 4.1 Electrons, nucleons and quarks 101
- 4.2 Elastic scattering 106
- 4.3 Inelastic *ep* scattering 111

5 Nuclear structure 117
- 5.1 Nuclear binding energies and nucleon forces 118
- 5.2 Nuclear models 126
- 5.3 Isospin and meson exchange 139
- 5.4 Radioactivity 147

6 Quantum Field Theory 161
- 6.1 Heisenberg's harmonic oscillator 161
- 6.2 Creation and annihilation operators 163
- 6.3 Interactions and fields 167
- 6.4 Relativistic quantum field theory 171
- 6.5 Bosons and forces 174

7 The Standard Model 176
- 7.1 Fermions and the generation puzzle 176
- 7.2 Bosons and the four forces 179

A major theme of 20th Century physics is that we are surrounded by substructure: what we see around us is built from smaller (often initially invisible) constituents and much of the diversity we see can be efficiently understood as consequences of the properties of these constituents. Furthermore this is a recursive process, with the constituents themselves often built from still-smaller pieces: matter is made of molecules; molecules are made of atoms; atoms made of nuclei and electrons; nuclei are built from nucleons (*i.e.* protons and neutrons); nucleons are made of quarks and gluons; and so on.

Subatomic physics is the part of this story starting with nuclei and continuing on to the smallest constituents known. We call particles 'elementary' if they have no substructure so far as we can tell, and at present the list of such particles contains around 20 entries. The theory of these particles and their mutual interactions is called the Standard Model and works extremely well (with a few noteworthy exceptions). But history teaches us that this classification of the elementary is at best provisional and may be changed in light of later evidence with finer resolution. These notes summarize the evidence for the present picture, as well as the flaws it is known to have, at a level appropriate for upper-year physics undergraduates. The reader is assumed to be familiar with non-relativistic quantum mechanics, electromagnetism and the rudiments of special relativity.

1 The story so far...

This section contains some preliminary background information needed to tell this story, and starts by summarizing the first indications that there might be an interesting story to tell.

1.1 Prequel: substructure and atoms

The evidence that many of the properties of macroscopic things are best understood if those things are regarded as being built of numerous much smaller constituents – atoms – started to accumulate convincingly in the 19th and early 20th Centuries. Partly this came about as the rules governing chemical reactions became clearer, with the emergence of a pattern of systematic properties for the elements, summarized by the periodic table (see Figure 9). Partly it emerged with the realization that the thermal properties of fluids (and thermodynamics in general) could be understood in terms of the random motion of their constituent atoms. It was clinched by the development of quantum mechanics and the ability this brought to compute the properties of simple atoms from first principles, including an understanding of the patterns of the periodic table.

Starting with Newton

In retrospect, the possibility that substructure could be useful was already implicit in the recursiveness of Newton's Laws. To see what this means, suppose that a macroscopic object, $\mathcal{O}$, is made up of a collection of N point-like atoms that mutually interact through forces $\mathbf{F}_{ij}$ (which describe the force acting on particle 'i' due to particle 'j'), with the atoms labelled by an index $i, j = 1, \cdots, N$. Then Newton's 2nd law for the motion of each atom is given by

$$\begin{aligned} m_1\,\ddot{\mathbf{x}}_1 &= \quad + \mathbf{F}_{12} + \mathbf{F}_{13} + \cdots + \mathbf{F}_{1N} + \mathbf{F}_{\text{ext}\,1} \\ m_2\,\ddot{\mathbf{x}}_2 &= \mathbf{F}_{21} \quad + \mathbf{F}_{23} + \cdots + \mathbf{F}_{2N} + \mathbf{F}_{\text{ext}\,2} \\ m_2\,\ddot{\mathbf{x}}_3 &= \mathbf{F}_{31} + \mathbf{F}_{32} + \quad \cdots + \mathbf{F}_{3N} + \mathbf{F}_{\text{ext}\,1} \\ \vdots \quad &= \\ m_N\,\ddot{\mathbf{x}}_N &= \mathbf{F}_{N1} + \mathbf{F}_{N2} + \mathbf{F}_{N3} + \cdots \quad + \mathbf{F}_{\text{ext}\,N}\,, \end{aligned} \tag{1.1}$$

where over-dots denote differentiation with respect to time – *i.e.* $\dot{\mathbf{x}} := \mathrm{d}\mathbf{x}/\mathrm{d}t$ and $\ddot{\mathbf{x}} := \mathrm{d}^2\mathbf{x}/\mathrm{d}t^2$ – while $\mathbf{F}_{\text{ext}\,i}$ denotes any external forces (*e.g.* attraction by the Earth's gravity *etc.*) acting on atom number 'i'.

The laws of motion for the entire macroscopic object must follow as consequences of eqs. (1.1), and at first sight it seems remarkable that any simple laws should be possible at all for macroscopic objects if this is so. A wonderful thing happens if all of these equations are added together, however, since then Newton's third law (which states that $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$ for all i and j) implies that all of the $\mathbf{F}_{ij}$ cancel in the sum, leaving

$$m_1\,\ddot{\mathbf{x}}_1 + m_2\,\ddot{\mathbf{x}}_2 + \cdots + m_N\,\ddot{\mathbf{x}}_N = \mathbf{F}_{\text{ext}\,1} + \cdots + \mathbf{F}_{\text{ext}\,N}\,. \tag{1.2}$$

This takes the same form as did Newton's law for each atom:

$$M\,\ddot{\mathbf{X}} = \mathbf{F}_{\text{ext}}\,, \tag{1.3}$$

with total mass and net external force given by

$$M := \sum_{i=1}^{N} m_i\,, \qquad \mathbf{F}_{\text{ext}} := \sum_{i=1}^{N} \mathbf{F}_{\text{ext}\,i}\,, \tag{1.4}$$

provided one defines

$$\mathbf{X} := \frac{1}{M}\sum_{i=1}^{N} m_i\,\mathbf{x}_i\,. \tag{1.5}$$

This shows that Newton's law applies in the same way to the entire macroscopic object provided the acceleration that appears in it is chosen to be the acceleration of the object's centre of mass — defined by (1.5).

Furthermore, this shows that Newton's 2nd law is *recursive* in the sense that it also applies equally well to various macroscopic subsets of macroscopic objects. For example suppose the

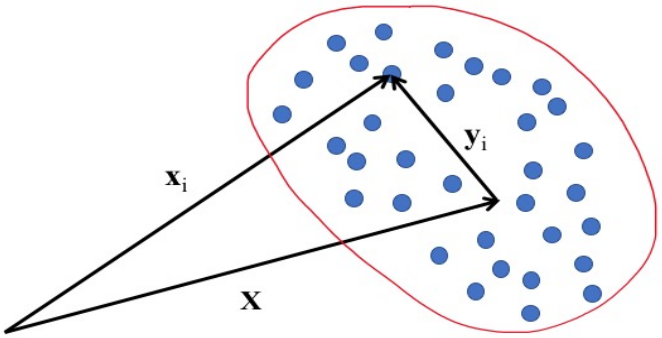


Figure 1. A sketch (not to scale) of atoms in a macroscopic object, illustrating the difference between the atomic position $\mathbf{x}_i$ and its position, $\mathbf{y}_i = \mathbf{x}_i - \mathbf{X}$, relative to the object's centre of mass, $\mathbf{X}$.

object described above can be regarded as the union of two pieces, denoted A and B, so $\mathcal{O} = A \cup B$. (Maybe the macroscopic object considered above was the Earth-Moon system and A is the Earth while B is the Moon.) Then all sums over i in the above argument can be broken up into sums separately over A and B:

$$M = \sum_{i=1}^{N} m_i = \sum_{i\in A} m_i + \sum_{i\in B} m_i =: M_A + M_B \,, \tag{1.6}$$

and similarly

$$\mathbf{F}_{\rm ext} = \sum_{i=1}^{N} \mathbf{F}_{{\rm ext}\,i} = \sum_{i\in A} \mathbf{F}_{{\rm ext}\,i} + \sum_{i\in B} \mathbf{F}_{{\rm ext}\,i} =: \mathbf{F}_{{\rm ext}\,A} + \mathbf{F}_{{\rm ext}\,B} \,. \tag{1.7}$$

So if we define

$$\mathbf{X}_A := \frac{1}{M_A} \sum_{i\in A} m_i\, \mathbf{x}_i \qquad \text{and} \qquad \mathbf{X}_B := \frac{1}{M_B} \sum_{i\in B} m_i\, \mathbf{x}_i \,, \tag{1.8}$$

then

$$M\,\ddot{\mathbf{X}} = \sum_{i=1}^{N} m_i\,\ddot{\mathbf{x}}_i = \sum_{i\in A} m_i\,\ddot{\mathbf{x}}_i + \sum_{i\in B} m_i\,\mathbf{x}_i = M_A\,\ddot{\mathbf{X}}_A + M_B\,\ddot{\mathbf{X}}_B \,, \tag{1.9}$$

where the last equality uses (1.8). Repeating the arguments leading to (1.3) separately for each of objects A and B then implies

$$M_A\,\ddot{\mathbf{X}}_A = \mathbf{F}_{{\rm ext}\,A} \qquad \text{and} \qquad M_B\,\ddot{\mathbf{X}}_A = \mathbf{F}_{{\rm ext}\,B} \,, \tag{1.10}$$

and (1.9) shows that these are consistent with (1.3).

Taken together, the above arguments show that the relationship between Newton's law for the whole system and Newton's law for its two subsystems is identical to the relationship derived earlier with Newton's laws for the N atoms, specialized to the case $N = 2$. (That is, it is conceptually as if each of A and B were themselves to be considered to be 'atoms'.)

This recursive nature of Newton's laws shows that the laws themselves cannot tell what the fundamental smallest objects are, and apply equally well at *all* levels of substructure. If tomorrow evidence were to emerge that all of our atoms in eq. (1.1) turn out to contain still-smaller teeny-weeny atoms, each of which themselves satisfy Newtons 2nd law, then nothing in the above arguments need change at all (provided we assume the position $\mathbf{x}_i$ to be an appropriately defined centre-of-mass coordinate).

Of course just because the internal forces cancel when Newton's 2nd law is summed over atoms doesn't mean that internal dynamics has no physical effect. For instance the average kinetic energy of motion of the constituent atoms relative to their centre of mass is a source of internal energy, since

$$E_{\rm kin} = \sum_{i=1}^{N} \frac{1}{2}\, m_i\, \dot{\mathbf{x}}_i^2 = \frac{1}{2}\, M\, \dot{\mathbf{X}}^2 + \sum_{i=1}^{N} \frac{1}{2}\, m_i\, \dot{\mathbf{y}}^2 =: \frac{1}{2}\, M\, \dot{\mathbf{X}}^2 + E_{\rm int}\,, \tag{1.11}$$

shows that it is an energy that survives even in the absence of motion of the object's overall centre of mass. In this expression $\mathbf{y}_i := \mathbf{x}_i - \mathbf{X}$ and no term linear in $\dot{\mathbf{y}}_i$ arises due to the easily proven identity

$$\sum_{i=1}^{N} m_i\, \mathbf{y}_i = 0\,, \tag{1.12}$$

that follows directly from the definition (1.5) of $\mathbf{X}$.

At this point $E_{\rm int}$ as defined above could equally well describe the energy of an overall rigid rotation of the macroscopic body, or the kinetic energy associated with altering its shape, or the random motion of its constituent atoms for an object whose macroscopic orientation and shape do not change. Indeed the discovery that the properties of this type of random internal energy reproduce those predicted by thermodynamics for large samples in equilibrium led to a deeper understanding of the statistical origins of temperature and heat in addition to providing part of the evidence supporting the existence of atoms.

Having the option to have substructure is not the same as there being compelling evidence for substructure's existence. This type of evidence mounted throughout the 19th century as chemistry became a quantitative science, whose reactions are well-described by the picture of molecular compounds swapping atoms. Rather than pursuing this story further (apart from a brief discussion of the Bohr atom, below) the rest of this introduction switches over to the line of thought that led to atoms themselves having substructure.

1.1.1 More than just atoms

The first discovery of a particle we now still regard as elementary occurred during the closing years of the 19th century. Many of these developments arose as unintended consequences of the discovery of *cathode rays* that are the result of applying a high voltage to a small amount of gas inside an otherwise evacuated tube. A very practical tube for producing these rays — called a Crookes tube — was developed as early as 1875. The presence of the cathode rays was inferred because they caused the gas in the tube (or any fluorescent material on the glass wall of the tube) to glow. (This is the phenomenon on which fluorescent lights and televisions were based until very recently).

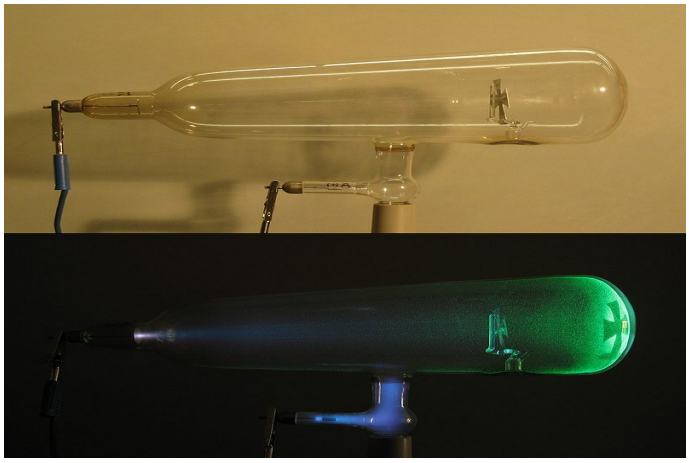


Figure 2. A Crookes tube with (bottom panel) and without (top panel) voltage applied, showing the fluorescense caused by the cathode rays when power is applied. The shadow of an obstacle shaped like a Maltese cross is visible and shows that the rays travel in straight lines. (Figure source: Wikipedia `https://en.wikipedia.org/wiki/Crookes_tube`).

We now know what happens with these tubes. The applied high voltage strips electrons from the gas atoms and the free electrons (*i.e.* the cathode rays) then are repelled by the negative cathode and flow towards the positive anode. Similarly, the leftover positive ions (or 'anode' rays) drift from the positive anode towards the cathode. The fluorescence is caused by collisions with these rays exciting gas atoms which emit light as they de-excite.

1.1.2 X rays

Many of the early workers with Crookes tubes (including Crookes himself, apparently) noticed that photographic plates became fogged up if they were near the tube when it was in use. Roentgen was the first to investigate in detail why this occurred, and by performing experiments with various objects between the tube and the plates found that images could be made of the dense parts within otherwise opaque objects. He determined that the plates

were fogged because they were being exposed to some new rays he called x-rays. By showing they moved in straight lines even in the presence of magnetic fields he showed these rays were electrically neutral.

We now know x-rays to be photons that are somewhat more energetic than visible or UV photons, and are emitted (like the fluorescent light) when excited electrons drop down into a state very close to the nucleus.

1.1.3 Radioactivity

In 1896 Becquerel, in Paris, ran experiments seeking to determine whether fluorescent materials could be made to emit x-rays through exposure to sunlight. To this end he took a good fluorescent compound, wrapped it in dark paper and placed it next to a photographic plate, intending to expose it to sunlight. Although his plans were thwarted when it was overcast in Paris that day, he nevertheless developed the plate and unexpectedly found that it had been exposed.

On further experimenting he determined that the exposure was due to the *spontaneous* emission of rays by the fluorescent material itself (rather than due to their fluorescing due to an applied voltage), since the photographic plates would become fogged regardless of whether or not the material was exposed to light or not. Furthermore, he found the rays responsible to be electrically charged since their path could be deflected by applying a magnetic field.

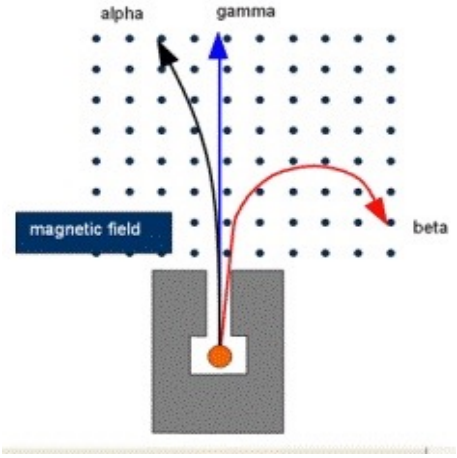


Figure 3. A schematic of how the paths of various radioactive rays respond to magnetic fields. The magnetic field points perpendicular to the page and is represented by the grid of dots. γ rays do not bend while α and β rays bend in opposite directions (because the particles involved have opposite-sign charges). β rays bend more strongly because their charge-to-mass ratio, q/m, is much larger. (Figure source: `http://www.particleadventure.org/radio_part.html`).

Over time it became clear that there are actually three different types of radioactive rays, called α, β and γ rays. Each responds differently to a magnetic field, with α rays behaving

like positively charged particles, β rays behaving like negatively charged particles and γ rays behaving like electrically neutral particles (and so are not deflected by magnetic fields). We now know what these 'rays' really are:

- α rays are made up of α particles, which are nuclei of ^{4}He — that is, two neutrons and two protons bound together. (Here the convention is that the left-superscript '4' of ^{4}He counts the total number of protons and neutrons present in the nucleus.) These types of Helium nuclei are particularly tightly bound and so very stable, and under certain circumstances can escape as a group from much larger nuclei. (In Becquerel's case it was the Uranium contained in his fluorescent compound that was responsible.)

- β rays consist of very energetic electrons that can be produced within the nucleus, usually through the decay of a neutron into a proton plus an electron (plus another particle, called an antineutrino — more about which later) through the reaction[1]

$$n \to p^+ + e^- + \overline{\nu}_e \,. \tag{1.13}$$

- γ rays are very energetic photons that differ from x-rays only by having more energy.

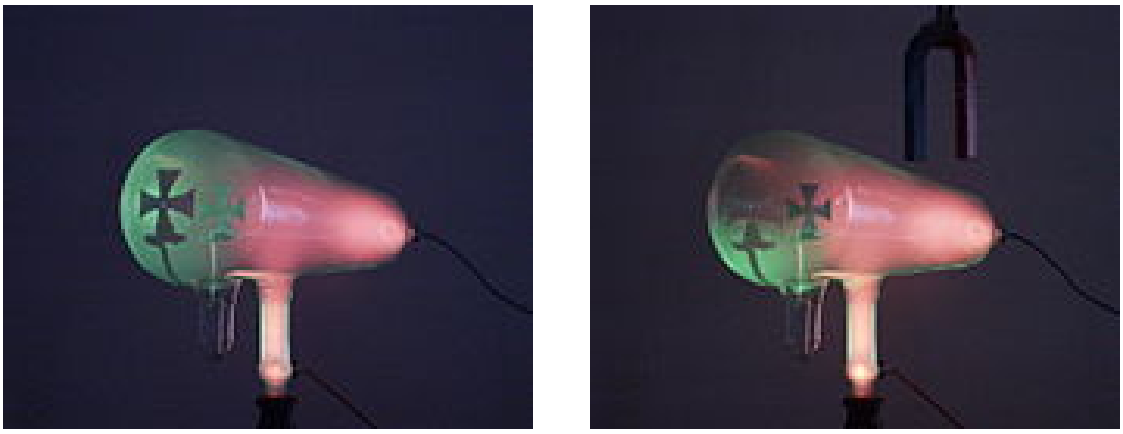
1.1.4 The electron

The nature of the cathode rays themselves was initially confused because of early experiments that incorrectly indicated that they were not deflected by magnetic fields (and so must be electrically neutral). (In retrospect these experiments were wrong because they were not able to get a good enough vacuum in the tube, and as a result ionization of the gas partially shorted out the voltage being applied to generate the electric field being used to test for a charge.) The situation was definitively settled in 1897 in experiments performed by J.J. Thompson, who was able to get a good enough vacuum in his apparatus to show that cathode rays are bent by a magnetic field, and in a direction that corresponds to being oppositely charged relative to α particles.

Furthermore, he used a clever combination of electric and magnetic fields to eliminate experimental unknowns and thereby pin down the charge-to-mass ratio of cathode rays. To do so Thompson produced cathode rays in the tube and deflected them by applying an electric field, as shown in Fig. 5.

He assumed the cathode ray was made of particles with mass m and charge q (this sounds uncontroversial to us, but at the time cathode rays were widely thought to be 'disturbances in the aether'), and that they are emitted from the cathode with speed v_0. By applying a

[1] When $\pm$ or $2\pm$ or $3\pm$ appears as a superscript for a particle label it represents the particle's electric charge in units of the proton charge, e. Hence p or p^+ is the proton and e^- is the electron. The superscript is often omitted for protons, as well as for neutral particles like neutrons (unless making a distinction with another particle with the same symbol: *e.g* the particle π^0 as opposed to π^+ or π^-).



(a) No magnetic field (b) Magnetic field applied

Figure 4. A Crookes tube with a no magnetic field applied (left panel) and with a magnetic field (right panel), showing how magnetic fields deflect cathode rays and so move the image of the Maltese cross on the fluorescent screen. (Figure source: Wikipedia `https://en.wikipedia.org/wiki/Crookes_tube`).

voltage between the plates marked P_1 and P_2 he arranged the particles to pass through a constant electric field, $\mathbf{E}$, and so to experience an electric force $\mathbf{F}_e = q\,\mathbf{E}$ and acceleration $\mathbf{a} = q\,\mathbf{E}/m$. The beam of particles is therefore deflected through an angle θ which Newton's Law gives to be

$$\tan\theta = \frac{v_y}{v_x} = \frac{a_y t}{v_0} = \frac{(qE/m)(\ell/v_0)}{v_0} = \frac{qE\ell}{mv_0^2}\,, \tag{1.14}$$

where the time spent between the plates is $t = \ell/v_0$, where ℓ is the easily measured length of the plates.

Although θ, E and ℓ could be measured, in order to infer q/m the problem was to determine the unknown initial speed, v_0. This he did by repeating the experiment with a magnetic field applied pointing perpendicular to the page, whose strength is adjusted to cancel the effects of the electric field so that the beam is not deflected at all. In order for this to be true the magnetic force, $\mathbf{F}_m = q\,\mathbf{v}\times\mathbf{B}$, must cancel the electric force so their magnitudes are related by $F_e = F_m$ and so $qE = qv_0B$. Solving then gives

$$v_0 = \frac{E}{B}\,, \tag{1.15}$$

and so using this in (1.14) allowed Thompson to determine the charge-to-mass ratio, q/m. The result was found to be much larger than that found for other rays, such as α radiation.

In retrospect this was the discovery of the first particle we still regard as being elementary. In particular, one sign that the cathode rays were something important was the fact that they are universal: they always have the same value of q/m regardless of the kind of dilute gas that is used in the tube. The same is *not* true of the 'anode' rays, which are the positively

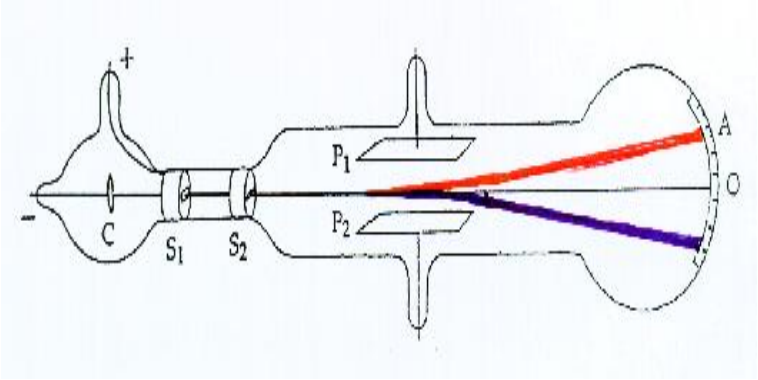


Figure 5. Thompson's apparatus for identifying the charge-to-mass ratio of cathode rays. High voltage is applied at the terminals marked + and − causing cathode rays to be emitted from the cathode marked C. The rays are collimated by passing through slits S_1 and S_2 to form a beam whose position can be seen on a fluorescent screen AO. Another voltage is applied between plates P_1 and P_2, forming an electric field in between that deflects the beam direction. Three sample beam trajectories are drawn, with the middle one corresponding to no voltage between P_1 and P_2 and the other two (coloured) trajectories corresponding to electric fields in opposite directions. (Figure source: `http://web.calstatela.edu/faculty/kaniol/f2000_lect_nuclphys/lect1/thompson_electron.htm`).

charged particles that are repelled by the anode and move towards the cathode. Anode rays are produced when Crookes tubes are set up with the voltages reversed, so that the source electrode at the left of above diagram is positively charged rather than negatively charged. When this is done the value of q/m found for these rays is much smaller than for cathode rays and, more importantly, has a value that depends on the precise gas used in the tube.

In retrospect what we now know is that applying a large enough voltage strips electrons from the atoms of the rarefied gas, after which the negative electrons are repelled by the cathode (and so are the cathode rays) while the positive ions are repelled by the anode (and so make up the anode rays). The fact that cathode rays always look the same is now understood because all atoms consist of electrons orbiting a nucleus, and although different elements have different nuclei (and so differing *numbers* of electrons in orbit) they are all built using the same type of electron.

1.1.5 Detection methods

A separate benefit of these various early experiments was the road they opened up to detecting the presence of these various types of new particles. After all, radioactivity and x-rays had always been around but went undetected because they were invisible to our senses. Their presence was eventually found due to their influence on atoms in other materials, such as the flourescent gas in a Crookes tube or the light-sensitive atoms in a photographic plate.

In retrospect detection occurred because collisions of the new particles with atoms in these materials excited the electrons within the atoms to higher-energy states and this was detected when these electrons then de-excited back to the ground state (often doing so by emitting light that was visible). Alternatively, more violent collisions between new particles and ambient atoms sometimes knocked electrons completely out of their atoms (in which case the rays were called 'ionizing radiation'). Such collisions also could be detected by observing the light emitted when the resulting ion subsequently recaptured another ambient electron, and cascaded down to its atomic ground state. Alternatively, the knocked-out electrons themselves can be detected by applying a voltage to the system and measuring the current caused as the knocked-out electrons drift towards the positive electrode.

These basic principles (measuring the photons and/or electrons produced by atomic excitation and/or ionization due to collisions with matter) remain the main techniques used for detecting elementary particles even now (though of course the detectors are much improved in efficiency relative to early days).

1.1.6 The nucleus

Having discovered the electron, and that electrons can be extracted from neutral atoms, Thompson was led to speculate about what the structure of the atom might be. In the absence of a better idea (and with the required tools like quantum mechanics not yet developed) he proposed the 'plum-pudding' model of the atom. In this model the atom is imagined to be a blob of positive charge (of unknown structure) within which electrons were uniformly distributed like the raisins in a pudding.

To test this model Rutherford performed an experiment in which he bombarded a thin gold foil with α particles that he obtained from the decay of a radioactive source. The idea was to watch how the alpha particles were scattered by the electrons and the positive charge within the atom, and use this to infer how they might be distributed. The apparatus is as illustrated in Fig. 6.

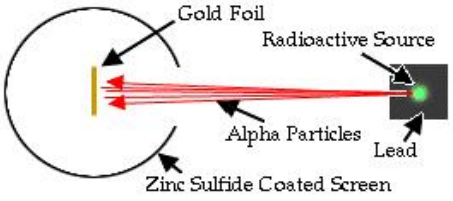
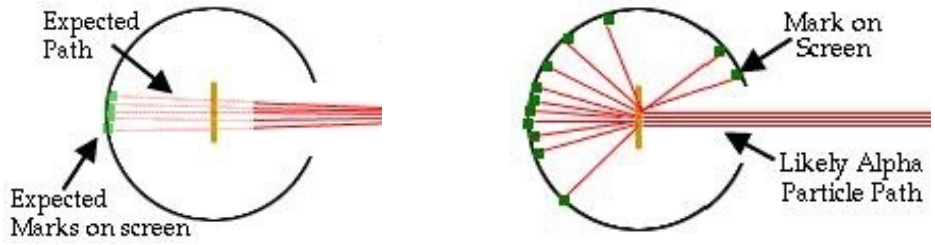


Figure 6. Schematic illustration of the apparatus with which Rutherford intended to probe the structure of the atom and test the plum-pudding model. The zinc-sulphide coated screen fluoresces when hit by α particles and so allows the direction of the scattered beam to be measured. (Figure source: Boston University `http://physics.bu.edu/cc104/chapters10and11.html`).

By this time the electric charge of the electron had been measured (through the Millikan oil-drop experiment of 1909) and so it was known that the electron had a charge equal in size to (but opposite in sign from) the charge, $q = e$, of the Hydrogen ion (what we now call the Hydrogen nucleus, or proton). The measurement of q/m for each then implied the electron was 1836 times lighter than a proton, and so that α particles were much heavier than electrons. As a result an α particle was expected only to scatter through a small angle, if at all, when encountering an electron. The same would also be true for scattering from a distributed positive charge distribution (as we see in detail in a later section), leading to the expectation that a plum-pudding atom would give the result illustrated in the left-hand panel of Fig. 7.

The experimental results therefore came as something of a surprise: while many alpha particles did only scatter through small angles some scattered much more strongly, even recoiling back into the same hemisphere from which they initially came (see the right-hand panel of Fig. 7). Furthermore, the measured probability of scattering as a function of the angle of the outgoing α-particle relative to its initial direction was consistent with that expected for scattering from the Coulomb potential of a point charge (more about this distribution below). Although Rutherford did not know the charge of a gold atom he thought it was likely to be roughly half its atomic weight, and so $Q \simeq 100\,e$. For this charge he could calculate the point of closest approach to the atom's central charge and so could put an upper limit on the size of the charge distribution to be $r_N \lesssim 10^{-14}$ m. This was already known to be much smaller than the radius, $r_A \sim 10^{-10}$ m, of the gold atom.



(a) Plum-pudding result (b) Nuclear result

Figure 7. Schematic illustration of the difference between what would be found in Rutherford's experiment if the plum-pudding model were true (left panel) and what was actually found (indicating the presence of a very compact source of positive charge – *i.e.* the nucleus. (Figure source: Boston University `http://physics.bu.edu/cc104/chapters10and11.html`).

1.1.7 The Nuclear Atom

The discovery of the nucleus set the stage for the development of atomic physics and quantum mechanics. Atoms became understood to consist of Z electrons (each with charge $-e$) orbiting

a nucleus with charge $Z\,e$ and mass M where $M = Am_0 + \delta$ is close to an integer times the atomic mass unit, m_0. Here δ can be negative but small: $|\delta| \ll m_0$ where m_0 is roughly the mass of a proton. (In practice m_0 is usually taken to be $\frac{1}{12}$ the mass of a Carbon nucleus, since this is better measured.) The positive integer Z is called the atom's atomic number or nuclear charge and the positive integer A is called its atomic mass number or atomic weight.

We now know the nucleus to be a bound state built out of a total of Z protons and $N = A - Z$ neutrons (both of whose masses are similar to m_0), and so the difference δ is to do with the binding energy that is responsible for holding the protons and neutrons together. Because protons each carry charge e and neutrons are neutral the nuclear charge is $Q = Z\,e$, and this determines the number of electrons needed to make the total atom neutral. Because chemical properties depend on the number of these electrons the number Z determines which element the atom corresponds to.

Although all atoms for any element share the same value of Z, they may differ in the number of neutrons present in their nucleus (and so differ also in their value for A). These different *isotopes* of an element are represented by ${}^{A}\mathrm{X}$ (where X is the symbol for the element — *e.g.* He for Helium or W for Tungsten, and the superscript A is the isotope's atomic weight). When the value of the nuclear charge is meant to be emphasized explicitly it can also be put in as a left-subscript,[2] as in ${}^{A}_{Z}\mathrm{X}$. For example ${}^{12}\mathrm{C}$ or ${}^{12}_{6}\mathrm{C}$ represents the most common isotope of Carbon whose nucleus contains 6 protons and 6 neutrons, while ${}^{14}\mathrm{C}$ or ${}^{14}_{6}\mathrm{C}$ represents a radioactive isotope of Carbon whose nucleus holds 6 protons but 8 neutrons.

For later purposes it is the success of the nuclear atom in explaining the chemical properties of elements, and of the frequency of the light emitted or absorbed during atomic transitions, that is particularly important. It had long been known that atoms only appeared to absorb or emit light with specific frequencies, with the pattern of allowed frequencies being characteristic of the element whose atom does the absorbing or emitting (see Fig. 8). The pattern for the frequency of light emitted by Hydrogen atoms in particular was characterized by a set of phenomenologically successful formulae whose physical origins were poorly understood.

The first steps at putting these formulae on a sound footing started with the conjecture by Neils Bohr that spectral measurements could be understood if the electrons in an atom were envisaged only to have specific allowed energies, labelled by a positive integer (now called the principal quantum number) $n = 1, 2, \cdots$. Bohr proposed that the frequencies of light emitted from Hydrogen would correspond to an electron made a 'quantum jump' (or transition) between two of the allowed energy levels, with the light's frequency, ω, satisfying

$$E_{\text{upper}} - E_{\text{lower}} = \hbar\,\omega\,, \tag{1.16}$$

[2] This notation leaves the right superscript free to indicate the ionic charge, should the nucleus not be surrounded by a full complement of electrons.

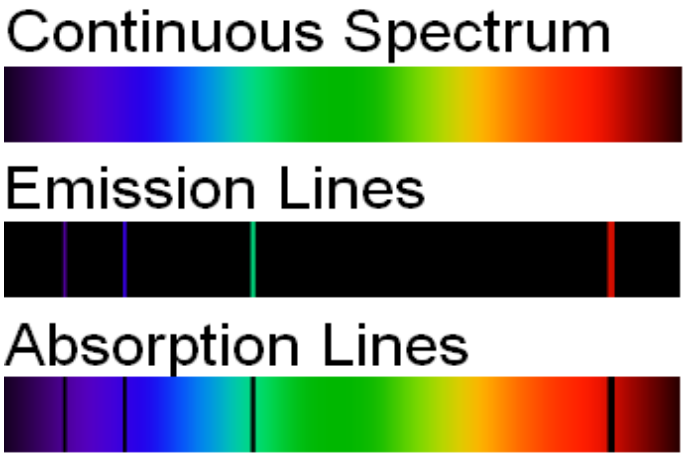


Figure 8. Top panel: the spectrum of light from a continuous source spread out using a prism. Middle panel: the spectrum of light emitted by a hot gas, showing the characteristic frequencies that are emitted by atomic transitions. Bottom panel: the absorption spectrum seen when light from a source emitting a continuous spectrum is viewed after being partially absorbed by passing through the same gas. Notice that emission and absorption occur at the same frequencies. (Figure source: Wikipedia Commons `https://commons.wikimedia.org/wiki/File:Spectral_lines_en.png`).

where $E_{\rm upper}$ and $E_{\rm lower}$ are the higher and lower electron energies. In particular the magnitudes of the allowed frequencies agreed with those observed by spectroscopists for Hydrogen if the quantized electron states had energies

$$E_n \simeq -\frac{\alpha^2}{2n^2}\, m_e c^2\,, \tag{1.17}$$

where m_e is the electron mass and $\alpha = e^2/(4\pi\hbar c) \simeq 1/137$ is the electromagnetic fine-structure constant and n is the electron's principal quantum number. The basic stability of atoms is then guaranteed by energy conservation since the minimum-energy ground state (the state with $n = 1$) has no lower-energy states into which it can decay.

A huge early success of quantum mechanics in its early days was its ability to derive eq. (1.17) for the allowed energies for electrons in a Hydrogen atom. Quantum mechanics obtains these energies as conditions for the existence of normalizable solutions to the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m_e}\,\nabla^2\psi - \frac{\alpha}{r}\,\psi = E\,\psi\,, \tag{1.18}$$

where $\psi(\mathbf{x}) = \psi(r,\theta,\varphi)$ is the wave function of an electronic moving within the Coulomb potential of the central nucleus. This equation also predicts the electron to be labelled by three integer quantum numbers: n, ℓ, ℓ_z. The 'principal' quantum number, $n = 1, 2, 3, \cdots$, is the positive integer anticipated by Bohr, while the 'angular-momentum' quantum number, ℓ,

takes values $\ell = 0, 1, 2, \cdots, n-1$ for every n and the 'magnetic' quantum number take values $\ell_z = -\ell, -\ell+1, \cdots, \ell-1, \ell$ for every ℓ.

Keeping in mind that an electron has spin $\frac{1}{2}$ and so has two spin states, these assignments mean that there are a total of

$$2 \sum_{\ell_z=-\ell}^{\ell} 1 = 2(2\ell+1) \tag{1.19}$$

states that share any specific value for the pair (n, ℓ), and so there are

$$\sum_{\ell=0}^{n-1} 2(2\ell+1) = 2n^2 \tag{1.20}$$

states sharing a specific value of n (and so also sharing a specific energy, E_n).

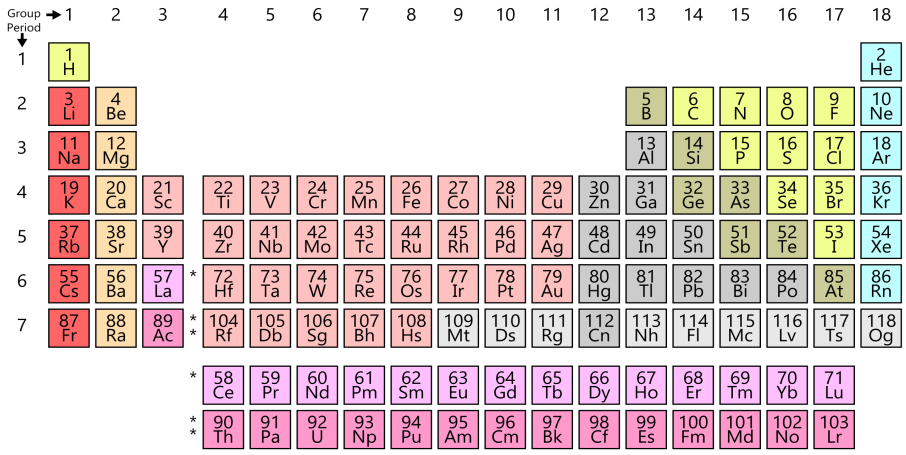


Figure 9. The periodic table, that groups elements into columns, all of whose members share similar chemical properties. The positive integer in each box is Z, the number of protons (or electrons) that are present in the neutral atoms of that corresponding element. (Figure source: Wikipedia `https://en.wikipedia.org/wiki/Periodic_table`).

The counting of levels sharing the same energy then goes a long way towards understanding the properties of the periodic table of the elements (shown in Fig. 9). The starting point is the observation that it is the properties of the least well-bound electron in an atom that control an element's chemical properties, since these are the electrons that are the least well bound to their own nucleus and so are most likely to be partly attracted to another nearby nucleus (thereby forming the interatomic bonds that drive the formation of molecules). To see how this helps understand the periodic table, neglect for illustrative purposes the repulsive Coulomb interactions among the electrons and focus only on the interactions of each electron with the nucleus. In this case each electron is described by a state satisfying (1.18), and so also has the binding energy given by (1.17).

Now comes the main point. If it is true that no two electrons can ever share the same state (a proposal, justified below, called the *Pauli Exclusion Principle*), then the lowest energy atomic configuration available for an atom containing Z electrons must fill the lowest energy levels available, starting with the ground state (with $n = 1$) and then moving up to the $n = 2$ state and so on until all Z electrons are placed. Eq. (1.20) predicts that only two can sit in the ground state (with $n = 1$), while 8 can sit in the $n = 2$ state once these are full, and 18 can have $n = 3$ and so on.

These resemble (with some caveats) the numbers of elements found in the top several rows of the periodic table. Only atoms with $Z = 1$ (Hydrogen) or $Z = 2$ (Helium) can have all electrons in the $n = 1$ ground state, and this corresponds to the periodic table's top row. Then the next 8 elements can have their outermost electron sit in an $n = 2$ state, corresponding to the table's second row. The outermost electron can be in an $n = 3$ state for the next 18 states and so on. This last number is not the right number for the table's third row but it is for its fourth. These discrepancies are understood in detail once the approximation of dropping inter-electron Coulomb repulsion is dropped. These interactions are more important for the lower rows of the periodic table rather than for the upper rows because as n gets larger it gets ever easier for repulsion to compete with the binding energy of the outermost electron

1.1.8 Bosons vs fermions

But what justifies the Pauli exclusion principle? Why shouldn't all electrons just sit all together in the ground state? The main reason for this goes back to whether electrons are bosons or fermions. But what are bosons and fermions?

The main point is that all elementary particles are identical. That means that any probability involving more than one particle, such as (for example) the joint probability, $P(\mathbf{x}_1, \mathbf{x}_2) = |\Psi(\mathbf{x}_1, \mathbf{x}_2)|^2$, to find the two particles to be located at the two position $\mathbf{x}_1$ and $\mathbf{x}_2$ must be symmetric: $P(\mathbf{x}_1, \mathbf{x}_2) = P(\mathbf{x}_2, \mathbf{x}_1)$ under any pairwise interchange of particles. This means in turn that the two-particle wavefunction must satisfy

$$\Psi(\mathbf{x}_1, \mathbf{x}_2) = \pm\Psi(\mathbf{x}_2, \mathbf{x}_1)\,. \tag{1.21}$$

Identical particles that satisfy (1.21) with upper (+) sign are called *bosons* while those that satisfy (1.21) with the lower sign (−) are called *fermions*. Both types of particles exist in Nature. It happens that all particles with integer spin $(0, 1, 2, \cdots)$ — such as photons, gravitons, Helium nuclei and the Higgs boson — are bosons while those with half-integer spin $(\frac{1}{2}, \frac{3}{2}, \cdots)$ — such as electrons, protons and neutrons — are fermions. (This connection between statistics and spin is not just an experimental fact; it is also a mathematical consequence of merging relativity and quantum mechanics, called the *spin-statistics theorem*.)

We can now see why no two fermions can ever be in the same quantum state. Suppose to this end that two identical non-interacting particles are independent of one another but interact with a common potential energy $V(\mathbf{x})$ (such as the gravitational field of the

Earth). In quantum mechanics the wave-function, $\Psi(\mathbf{x}_1, \mathbf{x}_2, t)$, for these particles satisfies the Schrödinger equation

$$i\,\frac{\partial \Psi}{\partial t} = -\frac{1}{2m}\Big(\nabla_1^2 + \nabla_2^2\Big)\Psi + \Big(V(\mathbf{x}_1) + V(\mathbf{x}_2)\Big)\Psi\,. \tag{1.22}$$

Because each term in this equation involves only $\mathbf{x}_1$ or $\mathbf{x}_2$ (and not both at once) it is always satisfied by the solution

$$\Psi(\mathbf{x}_1, \mathbf{x}_1, t) = \psi(\mathbf{x}_1, t)\,\tilde{\psi}(\mathbf{x}_2, t)\,, \tag{1.23}$$

where $\psi(\mathbf{x})$ and $\tilde{\psi}(\mathbf{x})$ are any two solutions to the single-particle Schrödinger equation

$$i\,\frac{\partial \psi}{\partial t} = -\frac{1}{2m}\nabla^2\psi + V(\mathbf{x})\psi\,. \tag{1.24}$$

A product solution like (1.23) is what one would expect if the two particles were statistically independent inasmuch as the joint probability distribution for finding them at a particular pair of positions also factorizes:

$$P(\mathbf{x}_1, \mathbf{x}_2, t) = |\Psi(\mathbf{x}_1, \mathbf{x}_2, t)|^2 = |\psi(\mathbf{x}_1, t)|^2 |\tilde{\psi}(\mathbf{x}_2, t)|^2 = P(\mathbf{x}_1, t)\tilde{P}(\mathbf{x}_2, t)\,, \tag{1.25}$$

as should statistically independent ensembles.

But condition (1.21) *forbids* choosing $\Psi(\mathbf{x}_1, \mathbf{x}_1, t) = \psi(\mathbf{x}_1, t)\,\tilde{\psi}(\mathbf{x}_2, t)$, and instead requires this to become at best

$$\Psi_\pm(\mathbf{x}_1, \mathbf{x}_2, t) = \frac{1}{\sqrt{2}}\Big[\psi(\mathbf{x}_1, t)\tilde{\psi}(\mathbf{x}_2, t) \pm \psi(\mathbf{x}_2, t)\tilde{\psi}(\mathbf{x}_1, t)\Big]\,, \tag{1.26}$$

where the upper (lower) sign applies for bosons (fermions). The two particles can never be in the same state because this last expression clearly shows that $\Psi_-(\mathbf{x}_1, \mathbf{x}_2, t)$ vanishes if we ever try to choose $\psi = \tilde{\psi}$.

1.1.9 The proton and neutron

Besides discovering the nucleus, Rutherford also pointed the way towards many experiments that followed since much would be learned about the structure of nuclei and their constituents by colliding them with other particles at high energies and studying what comes out. In the early days the particle beams used to probe the structure of atoms were α-particles coming from radioactive decays. Included amongst the discoveries arrived at in this way was the discovery of the particles we now know to be the constituents of atomic nuclei: protons and neutrons.

Protons were discovered to be nuclear constituents in experiments performed, again by Rutherford, in 1917 (but reported in 1919). In these Rutherford studied the *inelastic*[3] (and first ever man-made) nuclear reaction

$$\alpha + {}^{14}_{7}\mathrm{N} \to {}^{17}_{8}\mathrm{O} + p\,, \tag{1.27}$$

[3] A reaction is inelastic if the initial and final kinetic energies are not equal, so some internal energy is either absorbed or released.

by bombarding air with α particles. He determined that after the bombardment the air contained traces of Hydrogen that had not been previously present, and (rightly) concluded that he had knocked a Hydrogen nucleus out of one of the atomic nuclei. (It was the accumulation of traces of Helium outside of radioactive materials that similarly led to the conclusion that α particles were Helium nuclei.) The emerging proton was later seen more directly when the reaction took place within a cloud chamber, which is an early detector that allowed the direct measurement of the track of a quickly moving charged particle.

Since this discovery showed that nuclei could emit protons and β decays showed that nuclei could emit electrons it was natural to guess that nuclei were somehow built from protons and electrons. And because protons and electrons were known to carry equal but opposite electric charges, and protons were much more massive than electrons, the proper nuclear charge, Z, and mass number, A, could be achieved if nuclei could somehow be built from $A = Z + N$ protons plus N electrons, since this would ensure an atomic charge of Z and an atomic mass number of A.

Several things undermined this proposal in the end. First, nobody had a good explanation of the forces that would be required to bind protons and electrons into nuclei in this way. But this was not too daunting before the discovery of quantum mechanics because nobody then could understand how electrostatic attraction between electrons and protons could explain the orbits of electrons in an atom either. The discovery of quantum mechanics in the 1920s then resolved the problem of understanding how electrons move within atoms, but contrary to expectations it did not also in itself resolve the riddle of nuclear structure.

Quantum mechanics specifically undermined the idea that electrons and protons could bind within a nucleus in several ways. First, because both electrons and protons are fermions this model predicts that nuclei should satisfy Bose statistics whenever $N = A - Z$ is even, and should satisfy Fermi statistics whenever N is odd. In 1929 this ran into trouble once the vibrational spectrum of the Nitrogen molecule was measured. The ^{14}N nucleus has charge $Z = 7$ and mass $A = 14$ and so was expected to consist of 14 protons plus 7 electrons and therefore be a fermion. Yet observations instead showed that the statistical weights for the energy levels of the Nitrogen molecule required the wave-function to be symmetric under interchange of the Nitrogen nuclei: that is these nuclei behave as bosons. (More generally, observations show that nuclei are fermions whenever A is odd and are bosons whenever A is even.) Furthermore, it was also realized that the uncertainty principle requires the energy of an electron localized within something so small as a nucleus to be much higher than the energy associated with the electrons seen to emerge from nuclei in β decays.

> **Exercise 1.1:** Use position-momentum uncertainty relations, $\Delta x \Delta p \geq \hbar/2$, to estimate the lower limit to an electron's momentum if it is localized within a nucleus of size 1 fm $= 10^{-15}$ m. Given the relativistic energy-momentum relation, $E^2 = p^2c^2 + m^2c^4$, and electron mass ($mc^2 = 511$ keV) what is the electron's

kinetic energy ($E_{\rm kin} = E - mc^2$) corresponding to this momentum? How does this compare with the maximum electron energy (about 17 keV) seen in tritium β decay?

The ingredients required to properly understand the nucleus were finally in hand once the neutron was discovered in 1932. The discovery was just missed by Walther Bothe and Herbert Becker who found in 1931 that α particles bombarding Boron or Lithium produced some sort of radiation that was not bent by electric and magnetic fields. They therefore assumed these rays were γ rays, but this was made to seem doubtful because of the discovery by Irene Joliot-Curie and Frederic Joliot that these rays when impinging on paraffin (or other things containing Hydrogen) caused the production of very energetic protons. In 1932 James Chadwick, again by probing nuclei with α particles through the reaction [exercise 1]

$$\alpha + {}^{9}_{4}\mathrm{Be} \rightarrow {}^{12}_{6}\mathrm{O} + n\,, \tag{1.28}$$

showed that the new rays were electrically neutral particles whose mass was similar to that of a proton. Unlike γ rays, because of their mass neutrons carry enough momentum to knock a Hydrogen nucleus out of a sample when colliding with one, which explained the earlier observations with paraffin.

The discovery of the neutron allowed a number of things to be understood. Besides giving a better picture of the nucleus (more about which later), it opened the door to understanding β decay to be the result of neutrons within the nucleus decaying into protons and electrons (plus, it turned out, another undetected particle, the *neutrino* about which nothing was known at that time).

Neutrons also provided a new probe with which to bombard other nuclei, and they are particularly useful for this purpose (compared with protons or α particles) because their electrical neutrality means they are not repelled by the target nucleus' electric charge. Enrico Fermi found in 1934 that stable elements could be induced to become radioactive by bombarding them with neutrons, and by 1938 Otto Hahn, Lise Meitner and Fritz Strassmann discovered nuclear *fission* when they found that bombardment by neutrons could also split heavy nuclei into much much smaller pieces than happens through ordinary radioactivity.

1.2 Units and scales

For future purposes it is worth recording the units used throughout the rest of the notes.

1.2.1 Electron-Volts

Historically, the prominent role played by cathode rays made the *electron-Volt* a natural unit of energy:

An electron-Volt (or eV for short) is defined as the energy acquired by an electron falling through a voltage difference of one Volt, which implies 1 eV = 1.602176565× 10^{-19} J.

The usual metric conventions apply for multiples of this unit: 1 meV = 10^{-3} eV, 1 keV = 10^3 eV, 1 MeV = 10^6 eV, 1 GeV = 10^9 eV, 1 TeV = 10^{12} eV, and so on.

Electron volts continue to be used as natural units, but no longer because of cathode rays. Electron volts prove to be useful units for nuclear and particle physics because the rest mass of a proton (or a neutron) is, in energy units, 0.938 GeV (or 0.940 GeV). Because most matter is made of protons, neutrons and electrons, and because most of their mass comes from the rest mass of the protons and neutrons in their nuclei, this means that if you know the mass of an object in GeV then you also know how many nucleons it contains. For instance, knowing that there is $1.782661845(39) \times 10^{-25}$ kg per GeV/c^2 allows us to convert the mass of the Sun to

$$\begin{aligned} M_\odot &= 1.9885(2) \times 10^{30}\ \text{kg} \left(\frac{1.782661845(39) \times 10^{-25}\ \text{kg}}{\text{GeV}/c^2} \right)^{-1} \\ &\simeq 1.1155 \times 10^{55}\ \frac{\text{GeV}}{c^2}\,, \end{aligned} \tag{1.29}$$

and so see that the Sun contains roughly 10^{55} nucleons.

1.2.2 Fundamental units

Another convenient choice is to use units so that the main fundamental constants of nature are set to unity: *i.e.* choose units of length, time and temperature so that all three of the (reduced) Planck constant, speed of light and Boltzmann constant satisfy $\hbar = c = k_B = 1$. If this is done then it is no longer necessary to keep track of factors of these constants in expressions, which helps declutter formulae and makes it easier to see which variables are the important ones.

Once these units are used then we can measure any physical quantity in terms of a unit of length, say. (We could equally express everything in terms of a unit of time, or in terms of a unit of energy.) That is, if we say a time interval is measured in meters: $\Delta t = 3$ m, what we mean is that the time corresponds to how long it takes light to travel 3m, so there is an implicit unwritten factor of $c = 1$. The result in seconds can be found from $\Delta t = 3\,\text{m}/c = (3\,\text{m})/(3.0 \times 10^8\,\text{m/s}) = 10^{-8}$ s. The required power of c (or $\hbar$ or k_B) can be found using dimensional analysis. These units only make sense because everybody agrees on the values of c, $\hbar$ and k_B. The same argument allows mass to be written in units of energy where what is really meant by $m = 27$ J is $m = (27\,\text{J})/c^2 = (27\,\text{J})/(3.0 \times 10^8\,\text{m/s})^2 = 3.0 \times 10^{-15}$ kg.

Similarly the universal constant $\hbar = 1.1 \times 10^{-34}$ J-s allows energy to be converted to units of inverse seconds (or for time to be measured in units of inverse Joules). That is, we can arrange that $\hbar = 1$ (*i.e.* use natural units) if we measure energy in units of $\text{s}^{-1} = 1.1 \times 10^{-34}$

J. If someone tells us in natural units that $E = 80\,\mathrm{s}^{-1}$ then dimensional analysis tells us that there is an implicit, unwritten factor of $\hbar = 1$ and so to get the energy in Joules we write $E = 80\,\mathrm{s}^{-1} \times \hbar = (80\,\mathrm{s}^{-1})(1.1 \times 10^{-34}\,\mathrm{J\ s}) = 8.8 \times 10^{-33}$ J.

Because $\hbar$ has dimensions of (energy) $\times$ (time) it follows that $\hbar c = 3.3 \times 10^{-26}$ J-m has dimensions (energy) $\times$ (distance). This allows us to measure energy in inverse metres (or length in inverse Joules). For instance, the appropriate power of $\hbar c = 1$ that allows a statement like $E = 42\,\mathrm{m}^{-1}$ to make dimensional sense is $E = (42\,\mathrm{m}^{-1})\hbar c$ and so $E = (42\,\mathrm{m}^{-1})(3.3 \times 10^{-26}\,\mathrm{J\ m}) \simeq 1.4 \times 10^{-24}$ J.

Finally, we set $k_B = 1.4 \times 10^{-23}$ J/K to unity by agreeing to measure energy in degrees K or (more commonly) by measuring temperature in units of energy. In particular the choice made with fundamental units is to define the Joule as a unit of temperature so that 1 degree K equals 1.3807×10^{-23} J, since this ensures that $k_B = 1$. To convert temperature in J to temperature in K we just divide by k_B: *e.g.* $T = 280$ J in fundamental units really means the temperature in K is given by $T = (280\,\mathrm{J})/k_B = (280\,\mathrm{J})/(1.4 \times 10^{-23}\,\mathrm{J/K}) = 2 \times 10^{25}$ K.

Table 1. A selection of scales known to arise in nature

Measure in eV	Physical systems with these dimensions
10^{-32} eV	Hubble scale (inverse size/age of the universe as a whole)
10^{-23} eV	parsec (inverse distance to the nearest stars)
10^{-15} eV	light-second (inverse size of the Sun)
10^{-7} eV	inverse metre (everyday objects)
meV	energy levels of electrons shared by atoms in materials
eV	energy levels of outermost atomic levels
	(*e.g.* 1st Hydrogen excited state: 3.4 eV)
keV	energy levels of deeper atomic electrons for heavier elements
	(*e.g.* Hydrogen-like ground state with atomic no. Z: 13.6 Z^2 eV)
MeV	nuclear energy levels (1 - tens of MeV)
	electron rest mass ($m_e c^2 \simeq 0.5$ MeV)
GeV	nucleon rest mass ($m_p c^2 \simeq m_n c^2 \simeq 0.94$ GeV)
TeV	highest accelerator energies (LHC energy: 14 TeV)
10^{10} GeV	energies of the most energetic cosmic rays at Earth
$M_p c^2 \simeq 1.2 \times 10^{19}$ GeV	Planck mass: $M_p = (\hbar c/G_N)^{1/2}$

1.2.3 Hierarchies of scale

It is particularly useful to combine the above choices and so both adopt fundamental units and express all remaining quantities in dimensions that are a power of energy, with energy

measured in electron-Volts. This is very useful because the world around us is built from atoms and nuclei and so the scale of many phenomena are set by the size of the underlying physical properties like atomic or nuclear energy levels or particle rest masses. But these latter quantities have simple values expressed in terms of eV, so knowledge of a temperature or a distance in eV often also sheds light on the kind of physics (atomic, nuclear or other) involved (see Table 1).

For these purposes it is more useful to have $\hbar$ and k_B expressed using eV than with Joules. The corresponding present best numbers (and the value of c, for completeness) are

$$\hbar = 1.054571726(47) \times 10^{-34} \text{ J s} = 6.58211928(15) \times 10^{-22} \text{ MeV s}\,, \tag{1.30}$$

$$k_B = 1.3806488(13) \times 10^{-23} \text{ J / K} = 8.6173324(78) \times 10^{-5} \text{ eV / K}\,, \tag{1.31}$$

and

$$c = 2.99792458 \times 10^{8} \text{ m/s}\,, \tag{1.32}$$

so

$$\hbar c = 1.973269718(44) \times 10^{-7} \text{ eV m} = 197.3269718(44) \text{ MeV fm}\,, \tag{1.33}$$

where 1 femto-metre (or 1 Fermi or 1 fm) $= 10^{-15}$ m turns out to be close to the radius of a nucleus (which in turn is about 10^{-5} the radius of an atom). Roughly speaking these imply the following convenient rule-of-thumb conversions:

$$1 \text{ fm}^{-1} \sim 200 \text{ MeV} \sim (7 \times 10^{-24} \text{ s})^{-1}\,, \tag{1.34}$$

and

$$1\text{K} \sim 9 \times 10^{-5} \text{ eV}\,. \tag{1.35}$$

For convenience the Appendix provides several tables that convert between standard units for various quantities and their corresponding expressions in eV. When using these units it is useful to orient oneself by ordering several commonly occurring scales in physics as expressed in eV, as done in Table 1.

1.2.4 Cross-section units

Another possibly new unit used in these notes is the unit for scattering cross section (usually represented by the symbol σ). As is described in more detail in §2.3.1, a cross section is a useful measure of the likelihood of a reaction occurring when two different groups of particles are collided with one another. It has units of area, and can be loosely thought of as the area presented to a beam of particles by a target.

The default units for cross section would therefore seem to be m^2 or perhaps cm^2, but these are not so useful when considering targets the size of nuclei which are typically of order

1 fm – or 10^{-15} m– in radius. For this reason the standard conventional unit for cross section in nuclear and particle physics is the *barn* (or b) defined as

$$1\text{ b} = 10^{-24}\text{ cm}^2 = 10^{-28}\text{ m}^2 = (10\text{ fm})^2\,. \tag{1.36}$$

The usual metric prefixes apply, so 1 mb (or millibarn) is 10^{-27} cm^2, 1 μb (or microbarn) is 10^{-30} cm^2, 1 nb (or nanobarn) is 10^{-33} cm^2, 1 pb (or picobarn) is 10^{-36} cm^2 and 1 fb (or femtobarn) is 10^{-39} cm^2, and so on.

1.3 Lies, Damn Lies, and Measurement Errors

In reality any measurement has errors, and so any inference of physical properties (like particle masses, or energy levels, or decay rates) is uncertain inasmuch as repeated 'identical' measurements can return different values for the same quantity. This subsection provides a cartoon of some aspects of how these errors can be modelled probabilistically.[4]

Such uncertainties essentially make the result, x, of any particular measurement a random variable, with a characteristic probability distribution, $p(x)$, defined so that the probability that a measurement gives a result lying in a small interval x and $x + \mathrm{d}x$ is given by

$$P[x \in (x, x + \mathrm{d}x)] = p(x)\,\mathrm{d}x\,. \tag{1.37}$$

If $-\infty < x < \infty$ defines the range of all possible mutually exclusive outcomes of such a measurement, then the non-negative function $p(x)$ satisfies the normalization condition

$$P[x \in (-\infty, \infty)] = \int_{-\infty}^{\infty} \mathrm{d}x\; p(x) = 1\,. \tag{1.38}$$

Given such a random variable one can define expectation values, $\langle\, f(x)\,\rangle$, for functions, $f(x)$, as the sum over x of $f(x)$ weighted by the probability density $p(x)$:

$$\langle\, f(x)\,\rangle := \int_{-\infty}^{\infty} \mathrm{d}x\; f(x)\, p(x)\,. \tag{1.39}$$

Two special cases of this are the distribution's *mean*

$$\mu := \mathrm{M}(x) := \langle\, x\,\rangle := \int_{-\infty}^{\infty} \mathrm{d}x\; x\, p(x)\,. \tag{1.40}$$

and *variance*

$$\sigma^2 := \mathrm{V}(x) := \big\langle (x-\mu)^2 \big\rangle = \big\langle\, x^2 \big\rangle - \mu^2 = \int_{-\infty}^{\infty} \mathrm{d}x\; (x-\mu)^2\, p(x)\,. \tag{1.41}$$

[4] Quantum mechanics itself is intrinsically probabilistic and implies that repeated incompatible measurements – like alternating position and momentum measurements, for example – do not give precisely the same results on each iteration, and this happens even for ideal measurements. But that is not the worry here; the complaint is instead about the myriad of small uncontrolled influences – *i.e.* 'errors' – that reflect our imperfect understanding of the measurement process and whose influence can cause results to vary even for repeated classical measurements.

As defined, σ^2 is positive and its square-root, σ, is called the distribution's *standard deviation*. σ is a measure of the range of x values that are likely to be returned by repeated measurements.

Exercise 1.2: A random variable defined on the interval $0 \le x < \infty$ has an exponential probability distribution $p(x) = A\, e^{-x/a}$ for positive constants A and a. Compute the value of A required to ensure that $\int_0^\infty \mathrm{d}x\; p(x) = 1$. Compute the mean, μ, and standard deviation, σ, for this distribution.

Exercise 1.3: A random variable defined on the interval $0 \le x < \infty$ has a probability distribution

$$p(x) = \frac{1}{\Gamma(a) b^a}\; x^{a-1}\, e^{-x/b} \qquad \text{(Gamma distribution)} \tag{1.42}$$

for positive constants a and b. [Here $\Gamma(a+1) = a\,\Gamma(a)$ with $\Gamma(1) = 1$ is Euler's Gamma function.] Verify that $\int_0^\infty \mathrm{d}x\; p(x) = 1$. Show, for this distribution, that the mean is $\mu = ab$ and the variance is $\sigma^2 = ab^2$.

Error propagation

Sometimes one wishes to quantify how an error in one variable causes errors in other variables that are derived in terms of the first one. For example, for relativistic particles energy and mass are related to one another by $E = \sqrt{\mathbf{p}^2 + m^2}$ (using units for which $c = 1$), and one might wonder how much an uncertainty in E is caused by an error in the measurement of m.

Consider therefore a secondary quantity, $y = f(x)$, that is defined in terms of a random variable with known probability density $p(x)$. The mean and variance of y can be found using

$$\mu_y = \mathrm{M}(y) = \langle\, y \,\rangle = \langle\, f(x) \,\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\; f(x)\, p(x)\,, \tag{1.43}$$

and

$$\mathrm{V}(y) = \big\langle\, (y - \mu_y)^2 \,\big\rangle = \int_{-\infty}^{\infty} \mathrm{d}x\; \Big[f(x) - \mu_y \Big]^2 p(x)\,. \tag{1.44}$$

A notion of the uncertainty in y is given by its standard deviation, Σ, where $\Sigma^2 := \mathrm{V}(y)$.

There is a relatively simple estimate for the sizes of $\langle\, y \,\rangle$ and Σ in the event that $p(x)$ is sharply peaked around the mean, μ, (and the function $f(x)$ is not). In this case the above two integrals receive most of their support near $x = \mu$ and it becomes a good approximation to Taylor expand $f(x)$ around $x = \mu$,

$$f(x) \simeq f(\mu) + (x - \mu)\, f'(\mu) + \frac{1}{2}(x-\mu)^2 f''(\mu) + \cdots\,, \tag{1.45}$$

within the integrand, leading to the expression

$$\begin{aligned} \langle\, y \,\rangle \;&\simeq\; f(\mu) + f'(\mu) \int_{-\infty}^{\infty} \mathrm{d}x\; (x-\mu)\, p(x) + \frac{1}{2}\, f''(\mu) \int_{-\infty}^{\infty} \mathrm{d}x\; (x-\mu)^2\, p(x) + \cdots \\ &=\; f(\mu) + \frac{\sigma^2}{2}\, f''(\mu) + \cdots\,. \end{aligned} \tag{1.46}$$

This shows that the leading approximation is $\langle\, y \,\rangle = \langle\, f(x) \,\rangle \simeq f(\mu)$, and the error of dropping subdominant terms is of order $\frac{1}{2}\,\sigma^2\, f''(\mu)$, which is small when $f(x)$ is relatively slowly varying – compared to $p(x)$ – near $x = \mu$. Using $\langle\, y \,\rangle \simeq f(\mu)$ and (1.45) in the variance for y then gives the leading part

$$\begin{aligned}\Sigma^2 = \big\langle (y - \langle\, y \,\rangle)^2 \big\rangle &\simeq \int_{-\infty}^{\infty} \mathrm{d}x \, \Big[f(x) - f(\mu) \Big]^2 p(x) \simeq [f'(\mu)]^2 \int_{-\infty}^{\infty} \mathrm{d}x \, (x-\mu)^2 \, p(x) + \cdots \\ &\simeq \sigma^2 [f'(\mu)]^2 + \cdots \,, \end{aligned} \tag{1.47}$$

and so an estimate of the error in y generated by the error in x is given by

$$\Sigma \simeq \sigma\, f'(\mu) \,. \tag{1.48}$$

This is particularly simple for power-laws, $f(x) = Cx^n$ for constants C and n, since it implies that the *fractional* errors are then related by $\Sigma/\mu_y \simeq \sigma f'(\mu)/f(\mu) = n\sigma/\mu$.

Gaussian probability (Normal distribution)

Gaussian statistics are defined by the probability distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-(x-\mu)^2/(2\sigma^2)} \,, \tag{1.49}$$

also called a *Normal* distribution, $\mathcal{N}(\mu, \sigma^2)$. This is a particularly important type of probability distribution because (as seen below) combining a large number of independent random processes often leads to a normal distribution *even if the original processes do not themselves obey Gaussian statistics*. Although proving this assertion goes beyond the scope of this section, it is the underlying reason Gaussian statistics are chosen in the rest of the notes to be the probability density relevant to discussions of measurement error.

> **Exercise 1.4:** Prove that the parameter μ appearing in (1.49) is the distribution's mean — *i.e.* show that $\langle\, x \,\rangle = \mu$. Similarly prove that σ^2 is its variance, inasmuch as $\langle (x-\mu)^2 \rangle = \langle x^2 \rangle - \mu^2 = \sigma^2$.

In principle, $p(x)$ can be used to extract the probability that a measurement of x returns a result in any given range of values. For instance $\int_a^b \mathrm{d}x\, p(x) = P(a,b)$ is the probability of finding x in the interval $a \le x \le b$. Some useful probability statements that follow for the normal distribution are:

$$\begin{aligned} P[\mu - \sigma \le x \le \mu + \sigma] &\simeq 0.6826895 \\ P[\mu - 2\,\sigma \le x \le \mu + 2\,\sigma] &\simeq 0.9544997 \\ P[\mu - 3\,\sigma \le x \le \mu + 3\,\sigma] &\simeq 0.9973002 \\ P[\mu - 4\,\sigma \le x \le \mu + 4\,\sigma] &\simeq 0.9999367 \\ P[\mu - 5\,\sigma \le x \le \mu + 5\,\sigma] &\simeq 0.9999994 \end{aligned} \tag{1.50}$$

which give the probabilities of finding x to lie within 1, 2, 3, 4 and 5 standard deviations of the mean. A related (and similar) set of useful probabilities given by the Normal distribution are

$$\begin{aligned} P[\mu - 1.645\,\sigma \leq x \leq \mu + 1.645\,\sigma] &\simeq 0.90 \\ P[\mu - 1.960\,\sigma \leq x \leq \mu + 1.960\,\sigma] &\simeq 0.95 \\ P[\mu - 2.576\,\sigma \leq x \leq \mu + 2.576\,\sigma] &\simeq 0.99 \\ P[\mu - 3.290\,\sigma \leq x \leq \mu + 3.290\,\sigma] &\simeq 0.999 \end{aligned} \tag{1.51}$$

which give the range around the mean whose probability equals 90%, 95%, 99% and 99.9%.

> **Exercise 1.5:** Suppose that the probability for a runner to complete a marathon in time t is given by a Normal distribution, $\mathcal{N}(\mu, \sigma^2)$ (where in principle μ and σ take different values for each runner, and for the purposes of argument we allow t to run from $-\infty$ to ∞ even though in reality we know t must be positive). Suppose group A consists of runners, for all of whom $\mu_A = 270$ minutes and $\sigma_A = 15$ minutes, while all the runners from group B have $\mu_B = 270$ minutes with $\sigma_B = 20$ minutes (so both groups are equally fast on average but group B is more variable). Suppose 1000 members of group A and 1000 members of group B try out for the Olympic team by running a marathon for which only those who finish in less than 240 minutes are eligible to qualify. How many members from each of group A and group B should be expected to be eligible to qualify for the Olympic team? How many of each group are eligible if they are instead required to complete the marathon in 225 minutes?

Sampling and statistics

If errors are described by a Normal distribution, the value μ is the 'real' value the experimenter is trying to measure and σ is the size of the error forced on him/her by the measurement technique used. The goal is to design the experiment to minimize σ and thereby return measured values that are as close as possible to μ.

In practice experimenters usually do not know in advance what $p(x)$ is — or, for a Normal distribution, what μ and σ are. In principle this information can be extracted from the results of repeated measurements because $P(a, b)$ is related to (though not quite the same as) the fraction of times a sequence of identical measurements should find a result to lie between a and b. Strictly speaking, however, the fraction of repeated measurements that lie in (a, b) only precisely agrees with the probability $P(a, b)$ in the limit that the experiment is repeated an infinite number of times.

But any real experiment can only be done a finite number of times, producing a series of measurements x_i, with $i = 1, \cdots, N$. Given such a sample of measurements the experimenter

wants to infer $p(x)$ — or, for a Normal distribution the values of μ and σ. How can this be done? And how does the quality of this inference depend on things like the number of measurements, N, included in the sample? The idea is to choose an appropriate combination of the random variables, x_i, that has the property that it converges to the quantity of interest in the limit of infinite sample size, $N \to \infty$. Two examples of statistics[5] that can be estimators in this sense are the sample mean

$$\bar{x} := \frac{1}{N}\sum_{i=1}^{N} x_i \,, \tag{1.52}$$

and the sample variance

$$s^2 := \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2 \,. \tag{1.53}$$

These respectively prove to be estimators for the mean, μ, and variance, σ^2, of the underlying probability distribution.[6]

Any function of a random variable is itself a random variable, and this is true in particular for both $\bar{x}$ and s^2. That is, one can imagine drawing a sequence of N samples multiple times and finding that the mean and variance of these samples vary each time because the N-tuple of results, $\{x_1, \cdots, x_N\}$, also varies in each repetition. One can ask what the probability distribution is for $\bar{x}$ and s^2, given that each of the variables x_i is drawn from an ensemble with probability density $p(x)$.

There is a general answer to this question for the statistic $\bar{x}$, and it is such an important result that it has a name: the *central-limit theorem*. This states that as $N \to \infty$ the probability distribution for the variable $\bar{x}$ converges to a Normal distribution (1.49) with mean and variance given by:

$$\mathrm{M}(\bar{x}) = \frac{1}{N}\sum_{i=1}^{N} \mu_i = \mu \quad \text{and} \quad \mathrm{V}(\bar{x}) = \frac{1}{N^2}\sum_{i=1}^{N} \sigma_i^2 = \frac{\sigma^2}{N} \,. \tag{1.54}$$

What is remarkable about this statement is that it holds for *any* $p(x)$ provided this has finite mean and variance. In particular, for large N the mean of $\bar{x}$ converges to μ, with a spread around this mean of size

$$\frac{\sqrt{\mathrm{V}(\bar{x})}}{\mathrm{M}(\bar{x})} = \frac{\sqrt{\sum_i \sigma_i^2}}{\sum_i \mu_i} = \frac{\sigma}{\sqrt{N}\,\mu} = \frac{1}{\sqrt{N}}\left(\frac{\sqrt{\mathrm{V}(x)}}{\mathrm{M}(x)}\right) \,. \tag{1.55}$$

For samples of size N the distribution of $\bar{x}$ about its mean μ is smaller by a factor of $1/\sqrt{N}$ than is the spread about μ for each x_i separately.

[5] A 'statistic' is defined as a function of the random variables, x_i, that does not depend on any new parameters.

[6] Choosing the denominator in s^2 to be $N-1$ rather than N here turns out to be required to ensure that s^2 is an 'unbiased' estimator for the probability distribution's variance, σ.

When $p(x)$ itself is a Normal distribution then the probability distribution for s^2 is also known. The variable

$$z = \frac{(N-1)s^2}{\sigma^2} \qquad \text{then satisfies a } \textit{chi-squared} \text{ distribution} \quad \chi^2_{N-1}\,, \tag{1.56}$$

for $N-1$ degrees of freedom, where the χ^2_n probability distribution function is given by

$$p(\chi^2; n) = \frac{1}{2^{n/2}\Gamma(n/2)}\ \left(\chi^2\right)^{(n/2)-1}\ e^{-\chi^2/2} \qquad \text{for} \quad 0 \le \chi^2 < \infty\,. \tag{1.57}$$

Here $\Gamma(x)$ is Euler's gamma function, defined by $\Gamma(x+1) = \Gamma(x)$ and $\Gamma(1) = 1$, and so also satisfies $\Gamma(1/2) = \sqrt{\pi}$.

Furthermore if the x_i are statistically independent and Normally distributed then both $\bar{x}$ and s^2 are statistically independent. And if the x_i are independent and $\bar{x}$ and s^2 are independent then the x_i must be distributed Normally.

Exercise 1.6: Suppose a variable $-\infty < x < \infty$ is described by a probability distribution function $p(x)$. Suppose a sequence of N independent measurements are made that determine whether x is positive or not. The probability that x is positive in any one measurement is $\mathfrak{p} := \int_0^\infty p(x)\,\mathrm{d}x$. Show that the probability of obtaining precisely n results with $x > 0$ out of N independent trials is given by the *binomial distribution*:

$$P(n; N, \mathfrak{p}) = \frac{N!}{n!(N-n)!}\,\mathfrak{p}^n(1-\mathfrak{p})^{N-n}\,, \tag{1.58}$$

which satisfies $\sum_{n=0}^N P(n, N, \mathfrak{p}) = 1$ for all $0 \le \mathfrak{p} \le 1$. Show that the expected number of outcomes with $x > 0$ and the variance in this number are given by

$$\langle\, n \,\rangle := \sum_{n=0}^{N} nP(n; N, \mathfrak{p}) = N\mathfrak{p} \quad \text{and} \quad \left\langle n^2 \right\rangle - \langle\, n \,\rangle^2 = N\mathfrak{p}(1-\mathfrak{p})\,. \tag{1.59}$$

This shows how the fraction $\langle\, n \,\rangle/N$ of outcomes with $x > 0$ in a sample of measurements is an estimator for the probability $\mathfrak{p}$.

Exercise 1.7: Show that in the limit $N \to \infty$ and $\mathfrak{p} \to 0$ (with the product $\mathfrak{p}N = \mu$ held fixed) the binomial distribution of the previous problem becomes the *Poisson distribution*

$$P(n; N, \mathfrak{p}) \to P(n, \mu) := \frac{1}{n!}\,\mu^n e^{-\mu}\,, \tag{1.60}$$

for which $\langle\, n \,\rangle = \left\langle n^2 \right\rangle - \langle\, n \,\rangle^2 = \mu$. Useful for showing this is Stirling's formula for the factorial of a large number:

$$N! \simeq \sqrt{2\pi N}\ N^N e^{-N} \qquad (\text{for } N \gg 1)\,. \tag{1.61}$$

What is the variance of the distribution $P(n, \mu)$?

Exercise 1.8: Plot the Binomial distribution of Exercise 6 as a histogram for $N = 50$ and $\mathfrak{p} = 0.4$. On the same plot graph the Normal distribution with the same variance and mean and thereby see the Central Limit Theorem in action. A criterion for how big N must be in order to approximate the Binomial distribution with the Normal distribution is to ask for the 3-σ range of the Normal distribution, $(\mu - 3\,\sigma, \mu + 3\,\sigma)$, to lie within the range $(0, N)$, which requires

$$N > 9\left(\frac{1-\mathfrak{p}}{\mathfrak{p}}\right) \quad \text{and} \quad N > 9\left(\frac{\mathfrak{p}}{1-\mathfrak{p}}\right). \tag{1.62}$$

This (and the previous problem) shows in particular how convergence to a normal distribution can be slow when $\mathfrak{p}$ is very close to 0 or to 1.

Exercise 1.9: A particular reaction is predicted by a theory (the Standard Model, say) to occur N times in an accelerator experiment during a given period of time. But because the actual reactions are random events if the same experiment is repeated (each time for the same period of time) the number of events actually observed each time varies. Suppose the probability of actually observing n events in any one iteration of the experiment is given by the Poisson distribution $P(n, \mu)$ of Exercise 1.7 with mean $\mu = N$. With this distribution what is the probability of finding n to be larger than N by three standard deviations? What is the probability of finding n to be larger than N by five standard deviations?

Statistical and Systematic errors

With the above cartoon of statistics in our pocket, more can be said about what it means when an experimentalist says that a physical quantity, say E, is measured and the result is

$$E = E_0 \pm \Delta E\,. \tag{1.63}$$

If this is all that is said, usually what is being given is the mean and the standard deviation:

$$E_0 = \mathrm{M}(E) \qquad \text{and} \qquad (\Delta E)^2 = \mathrm{V}(E)\,, \tag{1.64}$$

of a sequence of measurements (or sample) $\{E_i\}$ where $i = 1, \cdots, N$, assumed to be randomly distributed with some probability distribution $p(E)$. Often (but not always) the distribution for this sample is assumed to be a normal distribution, in which case the variance is distributed by a χ^2 distribution.

Sometimes results are instead quoted as 'confidence intervals' (C.L.), with a statement like

$$E_0 - \Delta E < E < E_0 + \Delta E \quad \text{with } 95\% \text{ C.L.} \tag{1.65}$$

Translating this into a standard deviation is possible if the distribution $p(E)$ is known, with the conversion factors appropriate to a Normal distribution given in (1.50) and (1.51). For instance, inspection of (1.51) shows that 95% confidence level means that the quoted value for ΔE is related to standard deviation, σ, by $\Delta E \simeq 1.960\,\sigma$.

Of course the nature of uncertainty is that it is uncertain, and not all errors are reliably well-described as probabilistic. For instance sometimes there is simply a 'mistake' when analyzing an experiment, such as when there is a real physical effect that is present in reality but missing in the experiment's interpretation. This can happen both because of an honest-to-God mistake, or because the effect is not yet properly understood. In this case the error might not be equally biased in all directions, making it poorly described by a random sampling of a Gaussian (Normal) distribution. Such errors are often called 'systematic' errors (as opposed to 'statistical' errors for which probabilistic models are usually more suitable). When systematic errors are known to be present, careful experimenters sometimes quote errors of both types, as in: $E = E_0 \pm \Delta E_1$ (sys.)$\pm\Delta E_2$ (stat.).

What keeps things interesting is that one cannot always be sure at any given time that one understands all possible sources of error.

1.4 Relativistic kinematics

Table 1 shows that many energies of interest for this course are larger than the electron and proton rest energies, so for these it is important to use relativistic kinematics. This section is a refresher on those aspects of Special Relativity relevant to what follows.

1.4.1 Rotational invariance

From a practitioner's perspective Special Relativity is the statement that the laws of physics (*i.e.* of nature) are invariant under a symmetry, so before diving in it is worth first reviewing how things work for a similar symmetry: the invariance of nature's laws under rotation of an observer's reference frame.

Laws in physics (such as Newton's 2nd Law or the definition of kinetic energy)

$$\mathbf{F} = m\,\mathbf{a} \qquad \text{or} \qquad E_{\text{kin}} = \frac{m}{2}\,\mathbf{v}\cdot\mathbf{v}\,, \tag{1.66}$$

always come to us in the form vector = vector or scalar = scalar, but never have the form vector = scalar, say. There is a good reason for this, which is worth articulating explicitly.

In practice we usually use equations like $\mathbf{F} = m\,\mathbf{a}$ as a collection of component equations

$$F_x = m\,a_x\,, \qquad F_y = m\,a_y\,, \qquad F_z = m\,a_z\,, \tag{1.67}$$

where, for example, components like $F_i = \mathbf{e}_i\cdot\mathbf{F}$ (for $i = x, y, z$) denote the dot product between $\mathbf{F}$ and a basis of orthogonal unit vectors, $\mathbf{e}_i$, pointing along each of the three rectangular coordinate axes (and ditto for a_i and $\mathbf{e}_i\cdot\mathbf{a}$). We usually take for granted that the laws are

equally true regardless of the orientation in space used for the three basis vectors, $\mathbf{e}_i$. We can do so, but *only* because nature's laws don't have unusual forms like vector = scalar.

What is important is that both sides of equations like (1.66) transform in the same way under rotations, since this is what ensures component equations like (1.67) are the same[7] for any orthogonal basis vectors, $\mathbf{e}_i$. For instance, suppose we have two triads of orthonormal basis vectors, $\mathbf{e}_i$ and $\mathbf{e}'_i$, related to one another by rotation. Because rotation is linear (*i.e.* the rotation of zero is zero and the rotation of $\mathbf{a}$ and the rotation of $\mathbf{b}$ sum to the rotation of $\mathbf{a}+\mathbf{b}$) rotated basis vectors must be related by matrix multiplication

$$\begin{pmatrix} \mathbf{e}'_x \\ \mathbf{e}'_y \\ \mathbf{e}'_z \end{pmatrix} = \mathbf{R} \begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \\ \mathbf{e}_z \end{pmatrix} = \begin{pmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{pmatrix} \begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \\ \mathbf{e}_z \end{pmatrix} , \tag{1.68}$$

where R_{ij} are a collection of 9 real coefficients. We can write this relation more compactly in terms of the components of R using the notation

$$\mathbf{e}'_i = \sum_{j=x,y,z} R_{ij}\, \mathbf{e}_j = R_{ij}\, \mathbf{e}_j \,, \tag{1.69}$$

where the last equality introduces the *Einstein summation convention*, which suppresses the summation symbols by stating that any repeated subscript is implicitly meant to be summed over its entire range of values.

Given the matrix $\mathbf{R}$ the transformation of the components of any vector can be read off from the definitions:

$$F'_i = \mathbf{F}\cdot\mathbf{e}'_i = \sum_{j=x,y,z} R_{ij}\, \mathbf{F}\cdot\mathbf{e}_j = \sum_{j=x,y,z} R_{ij}\, F_j = R_{ij}\, F_j \,, \tag{1.70}$$

and similarly $a'_i = R_{ij}\, a_j$. In matrix form these become

$$\begin{pmatrix} F'_x \\ F'_y \\ F'_z \end{pmatrix} = \mathbf{R} \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} a'_x \\ a'_y \\ a'_z \end{pmatrix} = \mathbf{R} \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} , \tag{1.71}$$

with the same matrix $\mathbf{R}$, so the components of Newton's 2nd Law therefore become

$$\begin{pmatrix} F'_x \\ F'_y \\ F'_z \end{pmatrix} - \begin{pmatrix} m\,a'_x \\ m\,a'_y \\ m\,a'_z \end{pmatrix} = \mathbf{R} \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} - \mathbf{R} \begin{pmatrix} m\,a_x \\ m\,a_y \\ m\,a_z \end{pmatrix} = \mathbf{R} \left[\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} - \begin{pmatrix} m\,a_x \\ m\,a_y \\ m\,a_z \end{pmatrix} \right] . \tag{1.72}$$

This shows (because $\mathbf{R}$ is invertible) why the components of Newton's Law automatically apply in all rotated reference frames given that they apply in any one particular reference

[7] That is: if $F_i = m\,a_i$ in one frame this automatically ensures $F'_i = m\,a'_i$ for any rotated reference frame.

frame. What is important in the above argument is that every term in the equation transforms linearly, and each term transforms under rotations in exactly the same way, such as in (1.72).

It is also useful to be able to explicitly compute the coefficients R_{ij} for a specific rotation, and it is useful to know how many independent components of $\mathbf{R}$ there are. (In particular, does the matrix $\mathbf{R}$ above contain more than just the freedom to perform rotations?) For these purposes what is important is that all 9 of the components of $\mathbf{R}$ are not independent because equivalent observers also agree on the magnitude of any vector (and not just agree when a vector is zero, which is all something like (1.72) requires).

So we ask $\mathbf{R}$ not to change the orthonormality of the basis vectors, which is compactly expressed by $\mathbf{e}'_i \cdot \mathbf{e}'_j = \mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, with δ_{ij} denoting the Kronecker symbol whose defining properties are $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$. To see what this implies for R_{ij} take the dot product of (1.69) with itself, which shows

$$\delta_{ik} = \mathbf{e}'_i \cdot \mathbf{e}'_k = \sum_{j=x,y,z} \sum_{l=x,y,z} R_{ij} R_{kl}\, \mathbf{e}_j \cdot \mathbf{e}_l = \sum_{j=x,y,z} \sum_{l=x,y,z} R_{ij} R_{kl}\, \delta_{jl} = \sum_{j=x,y,z} R_{ij} R_{kj}\,, \quad (1.73)$$

or equivalently, with the Einstein summation convention,

$$\delta_{ik} = \mathbf{e}'_i \cdot \mathbf{e}'_k = R_{ij} R_{kl}\, \mathbf{e}_j \cdot \mathbf{e}_l = R_{ij} R_{kl}\, \delta_{jl} = R_{ij} R_{kj}\,. \quad (1.74)$$

Now the term on the far right-hand side is $R_{ij}R_{kj} = R_{ij}(R^T)_{jk} = (RR^T)_{ik}$ where $\mathbf{R}^T$ denotes the transpose of the matrix $\mathbf{R}$ and the last equality uses the definition of matrix multiplication. This shows that the matrix $\mathbf{R}$ is not an arbitrary one because it must satisfy the condition $\mathbf{R}\mathbf{R}^T = I$ where I is the unit matrix (whose components are δ_{ik}); that is to say $\mathbf{R}$ must be an orthogonal matrix.[8]

Since $(\mathbf{R}\mathbf{R}^T)^T = \mathbf{R}\mathbf{R}^T$ is a 3 by 3 symmetric matrix, it has 6 independent components and so the condition $\mathbf{R}\mathbf{R}^T = I$ imposes 6 conditions among the 9 components of the matrix $\mathbf{R}$. Using these 6 conditions to eliminate 6 of the components of $\mathbf{R}$ suggests $\mathbf{R}$ should contain a total of 3 free parameters, which turns out to be true. An arbitrary rotation matrix $\mathbf{R}$ turns out to be expressible in terms of products of a basic set of three independent rotations: a (clockwise) rotation about each of the three axes:

$$\mathbf{R}_x(\theta_x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_x & s_x \\ 0 & -s_x & c_x \end{pmatrix}, \quad \mathbf{R}_y(\theta_y) = \begin{pmatrix} c_y & 0 & s_y \\ 0 & 1 & 0 \\ -s_y & 0 & c_y \end{pmatrix}, \quad \mathbf{R}_z(\theta_z) = \begin{pmatrix} c_z & s_z & 0 \\ -s_z & c_z & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (1.75)$$

which for brevity writes $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$ for $i = x, y, z$ and the three angles, θ_i, are the three independent parameters in terms of which any 3-dimensional rotation can be described. It is straightforward to show that all three of these satisfy[9] $\mathbf{R}_i(-\theta_i) = [\mathbf{R}_i(\theta_i)]^T = [\mathbf{R}_i(\theta_i)]^{-1}$ for any θ_i, and so any matrix built from products of them must satisfy the defining property $\mathbf{R}\mathbf{R}^T = I$ for arbitrary θ_i.

[8] Because it involves the set of 3-by-3 orthogonal matrices this group of rotations is often called $O(3)$.

[9] Unusually, there is no Einstein summation convention used here.

1.4.2 Lorentz transformations

That familiar story about rotations sets up the following story about relativity. Special relativity states that the laws of nature are invariant under changes of reference frame in space *and* time amongst observers that move at constant velocity relative to one another, in such a way that all observers measure the same value for the speed of light. This condition can be framed in a very similar way in space-time as is done above for rotations in space.

To this end we use a basis of four unit vectors in space-time, three space unit vectors $\mathbf{e}_i$ as before plus one vector pointing in the time direction, $\mathbf{e}_t$. Rather than labelling space and time separately we collectively write the coordinates as

$$\{x^\mu\} = \{x^0, x^1, x^2, x^3\} = \{c\,t, x, y, z\} \tag{1.76}$$

using a Greek index $\mu = 0, 1, 2, 3$ and the convention that $\mu = 0$ corresponds to a time direction rather than a spatial one. (Very soon we adopt units with $c = 1$ in which case $x^0 = t$.) We wish to set up vectors in space-time (or 4-vectors), whose components — denoted V^μ — are obtained by taking dot products with a basis of vectors in space-time.

The dot product used in obtaining these components is the same as before in the spatial directions, but is modified in the time direction. This modification is chosen to ensure that the requirement that observers agree on the speed of light corresponds to the requirement they agree on the lengths of all 4-vectors in spacetime. To see what this means consider now a spherical light front that is emitted at some spatial position at a given time, $(t, \mathbf{x})$. After a small time interval, $\mathrm{d}t$, the position of the light front is given by the sphere of spatial radius $\mathrm{d}\mathbf{x}\cdot\mathrm{d}\mathbf{x} = c^2\mathrm{d}t^2$, so the set of points swept out by this light front (called the future light-cone of the emission event) satisfies

$$\begin{aligned} 0 = \mathrm{d}s^2 := -c^2\mathrm{d}t^2 + \mathrm{d}\mathbf{x}\cdot\mathrm{d}\mathbf{x} &= \begin{pmatrix} c\,\mathrm{d}t \\ \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix}^T \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} c\,\mathrm{d}t \\ \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} \\ &= \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} \mathrm{d}x^\mu\,\eta_{\mu\nu}\,\mathrm{d}x^\nu = \eta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu\,. \end{aligned} \tag{1.77}$$

The quantity $\mathrm{d}s^2$ defined here is called the invariant space-time interval, and special relativity requires all inertial observers must agree on its size.

The second line of (1.77) defines the components, $\eta_{\mu\nu}$, of the *Minkowski metric* for space-time. The very last equality uses the Einstein summation convention for the indices μ and ν to suppress the summation signs. Notice $\mathrm{d}s^2$ need not be positive: in particular $\mathrm{d}s^2 = 0$ for the surface of an expanding light wave, and intervals for which $\mathrm{d}s^2 = 0$ are therefore called *light-like* or *null*. Intervals with $\mathrm{d}s^2 > 0$ are called *space-like* because they include directions

separated only in space and not in time, while those with $\mathrm{d}s^2 < 0$ are called *time-like* because they include purely temporal intervals.

Special Relativity boils down to the requirement that inertial observers must be related by transformations that preserve the invariant interval, so its implications can be found in much the same way that rotations in the earlier section must preserve the magnitudes of vectors. Provided the laws of physics are expressed in terms of vectors for these transformations they will be the same for all such observers. To find what these transformations are we write a general linear transformation as $y^\mu = \sum_\nu \Lambda^\mu{}_\nu \, x^\nu = \Lambda^\mu{}_\nu \, x^\nu$, or in matrix form

$$\begin{pmatrix} y^0 \\ y^1 \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} . \tag{1.78}$$

Requiring the interval $\mathrm{d}s^2$ to be invariant for all 4-vectors requires the transformations $\Lambda^\mu{}_\nu$ must satisfy (switching permanently now to the Einstein summation convention)

$$\eta_{\mu\nu} = \Lambda^\lambda{}_\mu \, \eta_{\lambda\rho} \, \Lambda^\rho{}_\nu = (\Lambda^T \eta \, \Lambda)_{\mu\nu} \, . \tag{1.79}$$

Transformations that satisfy (1.79) are called *Lorentz transformations*, and because $\Lambda^T \eta \, \Lambda$ is a symmetric 4 by 4 matrix they impose 10 conditions on the 16 components of Λ, leaving a 6-parameter family of symmetries. But three of these parameters are old friends, since when Λ is restricted to act only in the spatial directions,

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & R^1{}_1 & R^1{}_2 & R^1{}_3 \\ 0 & R^2{}_1 & R^2{}_2 & R^2{}_3 \\ 0 & R^3{}_1 & R^3{}_2 & R^3{}_3 \end{pmatrix} , \tag{1.80}$$

condition (1.79) reduces to (1.74) and shows that the 3 by 3 submatrix $\mathbf{R}$ must be a spatial rotation.

The three new transformations are those that mix spatial directions with the time direction, and it is straightforward to verify that three independent solutions that satisfy (1.79) are given by the *boosts*

$$\Lambda_x(\beta_x) = \begin{pmatrix} \mathrm{ch}_x & \mathrm{sh}_x & 0 & 0 \\ \mathrm{sh}_x & \mathrm{ch}_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \quad \Lambda_y(\beta_y) = \begin{pmatrix} \mathrm{ch}_y & 0 & \mathrm{sh}_y & 0 \\ 0 & 1 & 0 & 0 \\ \mathrm{sh}_y & 0 & \mathrm{ch}_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} , \quad \Lambda_z(\beta_z) = \begin{pmatrix} \mathrm{ch}_z & 0 & 0 & \mathrm{sh}_z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \mathrm{sh}_z & 0 & 0 & \mathrm{ch}_z \end{pmatrix} , \tag{1.81}$$

where $\mathrm{ch}_i := \cosh \beta_i$ and $\mathrm{sh}_i := \sinh \beta_i$ for $i = x, y, z$.

What do these transformations mean physically? To determine this consider the action of Λ_x on the space-time coordinates: $y^\mu = \Lambda^\mu{}_\nu x^\nu$, where we drop the x subscript on Λ. Also writing $\beta_x = \beta$, this corresponds to the four component equations

$$y^0 = x^0 \cosh\beta + x^1 \sinh\beta\,, \quad y^1 = x^0 \sinh\beta + x^1 \cosh\beta\,, \quad y^2 = x^2 \quad \text{and} \quad y^3 = x^3\,, \tag{1.82}$$

or $c\,t' = c\,t\cosh\beta + x\sinh\beta$ and $x' = c\,t\sinh\beta + x\cosh\beta$ if $y^0 = c\,t'$, $x^0 = c\,t$, $y^1 = x'$ and $x^1 = x$ *etc.* These describe the coordinates of two observers that move relative to one another, as may be seen by asking how the curve $y^1 = y^2 = y^3 = 0$ (*i.e.* the origin of the spatial y^μ coordinates) looks in the x^μ coordinates. In particular, setting $y^1 = x' = 0$ implies x and t are related by

$$x = -c\,t\,\frac{\sinh\beta}{\cosh\beta} = -c\,t\,\tanh\beta \tag{1.83}$$

which shows the two observers move with constant relative speed, v, given by

$$\frac{v}{c} = \tanh\beta\,, \tag{1.84}$$

and so (using $\cosh^2\beta - \sinh^2\beta = 1$, which implies $\tanh^2\beta = 1 - 1/\cosh^2\beta$)

$$\cosh\beta = \frac{1}{\sqrt{1 - v^2/c^2}} =: \gamma \qquad \text{and} \sinh\beta = \frac{v/c}{\sqrt{1 - v^2/c^2}} = \frac{\gamma\, v}{c}\,, \tag{1.85}$$

where the first combination defines the quantity $\gamma(v)$. Eliminating β in favour of v in (1.82) reveals it to be the standard Lorentz transformation giving the time dilaton and the length contraction associated with motion along the x-axis:

$$t' = \gamma(t + xv/c^2)\,, \quad x' = \gamma(v\,t + x)\,, \quad y' = y \quad \text{and} \quad z' = z\,. \tag{1.86}$$

The transformations Λ_y and Λ_z similarly describe relative motion along the y and z axes. Boosts in an arbitrary direction can be built as appropriate products of Λ_x, Λ_y and Λ_z.

The quantity β related to v by (1.84) is called the *rapidity* of the relative motion and is useful because it transforms very simply when two successive boosts are performed in the same direction. That is, because matrix multiplication shows $\Lambda_x(\beta_1)\Lambda_x(\beta_2) = \Lambda_x(\beta_1 + \beta_2)$ the relativistic law for the addition of velocities is simply the addition of the two rapidities: $\beta_{12} = \beta_1 + \beta_2$. In terms of the speed, v, use of multiple-angle formulae for the hyperbolic trig functions shows the addition law for v is the familiar one

$$\begin{aligned}\frac{v_{12}}{c} &= \tanh\beta_{12} = \frac{\sinh(\beta_1+\beta_2)}{\cosh(\beta_1+\beta_2)} = \frac{\cosh\beta_2\sinh\beta_1 + \cosh\beta_1\sinh\beta_2}{\cosh\beta_1\cosh\beta_2 + \sinh\beta_1\sinh\beta_2} \\ &= \frac{\tanh\beta_1 + \tanh\beta_2}{1 + \tanh\beta_1\tanh\beta_2} = \frac{(v_1+v_2)/c}{1 + v_1 v_2/c^2}\,.\end{aligned} \tag{1.87}$$

In particular $v_1 < c$ and $v_2 < c$ imply $v_{12} < c$ and $v_{12} = c$ if either $v_1 = c$ or $v_2 = c$.

> **Exercise 1.10:** Calculate the relation between the coordinates $\{t', x', y', z'\}$ and $\{t, x, y, z\}$ obtained by first performing a boost in the x direction with speed v followed by a boost in the y direction with speed u.

1.4.3 Kinematic 4-vectors

Given this formulation of Special Relativity in terms of Lorentz transformations we see that the principle of Special Relativity amounts to the requirement that the laws of physics be Lorentz invariant. This will be automatic if these laws are expressed exclusively in terms of things that transform in the same way, that is laws of the form: 4-vector = 4-vector. Since laws of physics are cast in terms of position, velocity, momentum and acceleration, we next seek to identify the 4-vectors containing each of these.

Consider for these purposes a particle moving along some trajectory $\mathbf{r}(t)$ in space, not necessarily with constant velocity. Such a particle sweeps out a world-line in spacetime, and points along this world-line can be described by a one-parameter family of position 4-vectors

$$x^\mu(t) = \begin{pmatrix} c\,t \\ x(t) \\ y(t) \\ z(t) \end{pmatrix} \qquad \text{which has tangent} \qquad \frac{\mathrm{d}x^\mu}{\mathrm{d}t} = \begin{pmatrix} c \\ \mathrm{d}x/\mathrm{d}t \\ \mathrm{d}y/\mathrm{d}t \\ \mathrm{d}z/\mathrm{d}t \end{pmatrix} = \begin{pmatrix} c \\ \mathrm{d}\mathbf{x}/\mathrm{d}t \end{pmatrix} = \begin{pmatrix} c \\ \mathbf{v} \end{pmatrix}, \tag{1.88}$$

using the coordinates, t, x, y, z, of a specific observer. Although this has spatial components that agree with the particle's velocity, the problem with this definition is that it is not a 4-vector. That is, although any small displacement in spacetime, $\mathrm{d}x^\mu$, always transforms as a 4-vector, $\mathrm{d}x^{\mu\prime} = \Lambda^\mu{}_\nu\,\mathrm{d}x^\nu$, the time differential, $\mathrm{d}t$, is not a Lorentz-invariant measure of time and so $\mathrm{d}x^\mu/\mathrm{d}t$ does *not* transform as a 4-vector.

Much better instead to use arc-length measured along the particle world-line as the parameter, with distance defined using the invariant interval, $s(t)$, measured along the particle world-line. For any particle moving slower than the speed of light the infinitesimal interval measured along the world-line,

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}\mathbf{x}(t)\cdot\mathrm{d}\mathbf{x}(t) = -c^2\mathrm{d}t^2\left(1 - \frac{1}{c^2}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\cdot\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\right) = -c^2\mathrm{d}t^2\left(1 - \frac{\mathbf{v}\cdot\mathbf{v}}{c^2}\right), \tag{1.89}$$

is both Lorentz-invariant and always negative. So we define the infinitesimal *proper time* interval, $\mathrm{d}\tau$, along the particle world-line by:

$$\mathrm{d}\tau^2 := -\frac{\mathrm{d}s^2}{c^2} := \mathrm{d}t^2 - \frac{\mathrm{d}\mathbf{x}\cdot\mathrm{d}\mathbf{x}}{c^2} = \mathrm{d}t^2\left(1 - \frac{v^2}{c^2}\right), \tag{1.90}$$

where $v^2 := \mathbf{v}\cdot\mathbf{v}$ as usual. This gets its name because it agrees with the time interval, $\mathrm{d}t$, measured by a clock that is instantaneously in the rest frame of a particle; (*i.e.* one for which $\mathrm{d}\mathbf{x} = 0$ in the interval $\mathrm{d}t$). Notice that (1.90) implies such a clock evolves in the way required by time-dilation relative to an observer at rest because a proper-time interval, $\mathrm{d}\tau$, is related to the interval, $\mathrm{d}t$, of the observer at rest by[10]

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\sqrt{1 - v^2/c^2}} = \gamma\,. \tag{1.91}$$

[10] We choose the positive root here so that $\mathrm{d}\tau$ is positive whenever $\mathrm{d}t$ is.

This suggests defining the velocity 4-vector, or 4-*velocity*, u^μ, by

$$u^\mu := \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \frac{\mathrm{d}x^\mu}{\mathrm{d}t}\,\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\sqrt{1-v^2/c^2}}\begin{pmatrix} c \\ \mathbf{v}\end{pmatrix} = \begin{pmatrix} \gamma\, c \\ \gamma\,\mathbf{v}\end{pmatrix}, \tag{1.92}$$

and this indeed transforms like a 4-vector, $u^\mu \to \Lambda^\mu{}_\nu\, u^\nu$, because of the transformation rule $\mathrm{d}x^\mu \to \Lambda^\mu{}_\nu\, \mathrm{d}x^\nu$ and the invariance of the interval $\mathrm{d}\tau$. Notice that this definition implies a particle's 4-velocity always has the following invariant norm:

$$\eta_{\mu\nu}\, u^\mu u^\nu = -(u^0)^2 + \mathbf{u}\cdot\mathbf{u} = -\gamma^2\left(c^2 - \mathbf{v}\cdot\mathbf{v}\right) = -c^2\,. \tag{1.93}$$

The particle's 4-momentum is defined as being proportional to the 4-velocity:

$$p^\mu := m\, u^\mu = \begin{pmatrix} \gamma\, mc \\ \gamma\, m\mathbf{v}\end{pmatrix} = \begin{pmatrix} E/c \\ \mathbf{p}\end{pmatrix}, \tag{1.94}$$

where we use the standard definitions for the relativistic momentum and kinetic energy:

$$E = \gamma\, mc^2 \qquad \text{and} \qquad \mathbf{p} = \gamma\, m\mathbf{v}\,. \tag{1.95}$$

Eq. (1.93) and the definition $p^\mu = m\, u^\mu$ implies E and $\mathbf{p}$ are related to one another by

$$\eta_{\mu\nu} p^\mu p^\nu = -(E/c)^2 + \mathbf{p}^2 = -(mc)^2\,, \tag{1.96}$$

which implies the standard energy-momentum relation

$$E = \sqrt{\mathbf{p}^2c^2 + (mc^2)^2}\,, \tag{1.97}$$

that for $|\mathbf{p}c| \ll mc^2$ approximately reproduces the nonrelativistic expression $E \simeq mc^2 + (\mathbf{p}^2/2m) + \mathcal{O}[(\mathbf{p}c)^4/(mc^2)^3]$. This allows m to be interpreted as the particle's rest-mass. (From here on we use the words rest-mass and mass interchangeably.) It also can be rewritten in the following two useful results giving γ and $\mathbf{v}$ in terms of E and $\mathbf{p}$:

$$\gamma = \frac{E}{mc^2} \qquad \text{and} \qquad \frac{\mathbf{v}}{c} = \frac{\mathbf{p}\,c}{E}\,. \tag{1.98}$$

Exercise 1.11: As an example of the utility of knowing that quantities like p^μ and u^μ transform as 4-vectors under Lorentz transformations, prove that

$$E = -u_\mu\, p^\mu = -\eta_{\mu\nu}\, u^\mu\, p^\nu\,, \tag{1.99}$$

is Lorentz-invariant and gives the energy of a particle with 4-momentum p^μ as seen by an observer with 4-velocity u^μ. (*Hint:* use that E is the same in all frames plus the information that $u^\mu = \{c, 0, 0, 0\}$ in the rest-frame of the observer for which u^μ is the 4-velocity.

In the absence of an external force Einstein's generalization of Newton's 2nd Law states that p^μ is strictly conserved, and this encodes both conservation of kinetic energy and conservation of momentum for a free particle.

Although inertial observers must move relative to one another with constant velocity, nothing in special relativity stops you from considering how these observers describe the trajectory of particles that accelerate. For instance, consider a trajectory describing a particle that accelerates along the x axis from rest at $x = 0$, until its speed reaches $v = v_{\max}$ at which point it then decelerates back to rest a distance ℓ away and then returns to $x = 0$, again at rest according to the specific rule

$$x^\mu(t) = \Big\{ ct, x(t), y(t), z(t) \Big\} = \left\{ ct, \ell \sin^2\left(\frac{v_{\max} t}{\ell} \right), 0, 0 \right\} . \tag{1.100}$$

Here the inertial observer's time, t, is used to label points on the curve, with $0 \leq t \leq T = \pi\ell / v_{\max}$ describing the entire round trip. The turning point at $x = \ell$ is achieved at $t = \frac{1}{2} T$, and because the instantaneous particle speed seen by the inertial observer is

$$v(t) = \frac{\mathrm{d}x}{\mathrm{d}t} = v_{\max} \sin\left(\frac{2 v_{\max} t}{\ell} \right) , \tag{1.101}$$

the maximum speed on the outbound leg takes place at $t = (\pi\ell / 4 v_{\max}) = \frac{1}{4} T$.

The proper time measured by a clock riding with the particle along such a trajectory is

$$\mathrm{d}\tau^2 = -\frac{\mathrm{d}s^2}{c^2} = -\eta_{\mu\nu} \frac{\mathrm{d}x^\mu(t) \mathrm{d}x^\nu(t)}{c^2} = \left[1 - \frac{v^2}{c^2}(t) \right] \mathrm{d}t^2 , \tag{1.102}$$

and so the instantaneous 4-velocity and 4-acceleration are

$$\begin{aligned} u^\mu &= \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \frac{\mathrm{d}t}{\mathrm{d}\tau} \frac{\mathrm{d}x^\mu}{\mathrm{d}t} = \frac{1}{\sqrt{1 - v^2(t)/c^2}} \Big\{ c, v(t), 0, 0 \Big\} \\ \text{and} \quad a^\mu &:= \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} = \frac{\mathrm{d}t}{\mathrm{d}\tau} \frac{\mathrm{d}u^\mu}{\mathrm{d}t} = \frac{\mathrm{d}v/\mathrm{d}t}{[1 - v^2(t)/c^2]^2} \Big\{ v(t)/c, 1, 0, 0 \Big\} , \end{aligned} \tag{1.103}$$

with

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{2 v_{\max}^2}{\ell} \cos\left(\frac{2 v_{\max} t}{\ell} \right) . \tag{1.104}$$

In relativistic Newtonian mechanics the force responsible for this motion is described by a 4-vector, $F^\mu = m\, a^\mu$, and all inertial observers must agree on the *proper acceleration* given by the Lorentz-invariant definition

$$a^2 := \eta_{\mu\nu}\, a^\mu a^\nu = a_\mu a^\mu = \frac{1}{[1 - v^2(t)/c^2]^3} \left(\frac{\mathrm{d}v}{\mathrm{d}t} \right)^2 . \tag{1.105}$$

Exercise 1.12: Compute the proper time, 4-velocity, 4-momentum and 4-acceleration for the following trajectories: (a) constant proper acceleration along the z axis, $x^\mu(u) = \{\ell \sinh(\alpha u), 0, 0, \ell \cosh(\alpha u)\}$, and ($b$) uniform circular motion in the x-y plane, $x^\mu(u) = \{ct, d\cos(\omega t), d\sin(\omega t), 0\}$. What is the physical interpretation of the parameters ℓ, α, d and ω used in these trajectories?

Exercise 1.13: Suppose a family of light rays having frequency ω is sent parallel to the x-y plane at an angle θ to the x axis, and so has 4-momentum $k^\mu = \{\hbar\omega_\star, (\hbar\omega_\star/c)\cos\theta, (\hbar\omega_\star/c)\sin\theta, 0\}$. Show that this satisfies $k_\mu k^\mu = 0$, as it must if it is tangent to the trajectory of a light ray. Use the relation $E = \hbar\omega$ and $E = -\eta_{\mu\nu}\, u^\mu k^\nu$ to evaluate the frequency of the photons that is measured by observers moving along the accelerated trajectories in the previous exercise.

2 Calculational tools I

Since much of what we know about subnuclear physics comes from studying collisions and decays, in this section we collect some useful tools for analyzing these types of processes.

Measurement of a decay or scattering rate carries two kinds of information: information following from conservation laws and information that goes beyond simple conservation. Consequences of conservation laws have the advantage of being very robust: their validity does not depend on the details of the forces involved so long as these conserve the things of interest (*e.g.* energy, momentum, angular momentum, electric charge *etc*). It is the information that does *not* follow simply from conservation that is most informative about the nature of the interactions that are responsible for a decay or a scattering event.

2.1 Conserved quantities

There are a number of quantities that are known whose conservation, or approximate conservation, plays an important role in constraining scattering and decay processes. All experiments performed to date are consistent with the following quantities being exactly conserved:

- *Energy - Momentum*, p^μ, is believed to be exactly conserved, and the conservation of the four components of 4-momentum contain what would be (for Newtonian physics) the separate conservation laws of energy and momentum.

- *Angular Momentum* is believed to be exactly conserved, and so each particle is assigned a value for its total angular momentum, J, with $J = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$, and contains $2J+1$ states corresponding to the allowed values of the 3rd component of angular momentum, $J_3 = -J, -J+1, \cdots, J-1, J$. The rules of combining angular momenta then restrict (for example) the spins and orbital angular momenta that can appear among the daughter products in terms of the spin of a decaying particle.

- *Electric charge*, Q, is also believed to be exactly conserved, and all particles ever seen experimentally have an integer multiple of the proton charge, e, though there is nothing in principle[11] that requires this (and so would forbid having fractional charges).

- *Baryon number*, B, appears to be conserved in practice, though the best theories at present do not require this conservation to be exact. Protons and neutrons (plus other particles, called baryons) each carry baryon number $B = +1$ and their anti-particles (the antiproton and antineutron) carry baryon number $B = -1$. Other particles mentioned to this point, such as electrons, have $B = 0$.

- *Lepton number*, L appears to be conserved in practice, but need not be exactly conserved in principle. Of the particles discussed to this point, electrons, muons and neutrinos all carry lepton number $L = +1$, and their antiparticles carry lepton number $L = -1$. All others (in particular protons and neutrons) carry $L = 0$.

There are also a number of quantities that appear to be approximately conserved, in the sense that they are conserved by *almost* all of the interactions in nature, and so are for most purposes useful conservation laws. But they are broken by small detectable amounts in a few specific situations. The most important of these for the purposes of the first half of these notes are

- *Electron number*, L_e, is defined so that the electron, e^-, and electron neutrino, ν_e, carry $L_e = +1$ while their antiparticles, e^+ and $\overline{\nu}_e$, carry $L_e = -1$. All other particles carry $L_e = 0$.

- *Muon number*, L_μ, is defined in a similar way as electron number, but for muons. Muons and muon neutrinos, μ^- and ν_μ, carry $L_\mu = +1$ while their antiparticles, μ^+ and $\overline{\nu}_\mu$, carry $L_\mu = -1$. All other particles carry $L_\mu = 0$.

- *Isospin*, T and T_3, are two approximately conserved labels that particles carry that are very much like the labels J and J_3 for angular momentum, with $T = 0, \frac{1}{2}, 1, \frac{3}{2}, \cdots$ and $T_3 = -T, -T+1, \cdots, T-1, T$. Unlike for angular momentum the states corresponding to different labels for T_3 are different particles (rather than just different 'spin' states of the same particle). We shall see how the approximate conservation of T and T_3 expresses how nuclear forces seem to treat several types of particles (notably protons and neutrons) in almost exactly the same way.

[11] More precisely: within the framework of the Standard Model combined with General Relativity the condition that there be no gauge anomalies (including mixed gravitational anomalies) actually does determine all ratios of electric charge. But if one broadens the framework to include more, hitherto undetected, particles the same need not remain true. Here an 'anomaly' is when the classical conservation of a charge fails to survive quantization (which can sometimes happen), and a 'gauge anomaly' is when such an anomaly occurs for a charge (like electric charge) that is the source of a long-range force. Gauge anomalies are believed not to arise in sensible theories since they violate either the unitarity of quantum mechanics or Lorentz-invariance.

As discussed in more detail below, before the discovery of neutrino oscillations in the 1990s both L_e and L_μ were also believed to be effectively[12] exact conservation laws. Table 2 gives a table of these quantum numbers for the most commonly occurring particles:

Table 2. Charge assignments for a selection of common particles

Particle (symbol)	rest mass	J	Q/e	B	L	L_e	L_μ	T	T_3
photon (γ)	0	1	0	0	0	0	0	0	0
proton (p)	938 MeV	$\frac{1}{2}$	$+1$	$+1$	0	0	0	$\frac{1}{2}$	$+\frac{1}{2}$
antiproton ($\overline{p}$)	938 MeV	$\frac{1}{2}$	-1	-1	0	0	0	$\frac{1}{2}$	$-\frac{1}{2}$
neutron (n)	940 MeV	$\frac{1}{2}$	0	$+1$	0	0	0	$\frac{1}{2}$	$-\frac{1}{2}$
antineutron ($\overline{n}$)	940 MeV	$\frac{1}{2}$	0	-1	0	0	0	$\frac{1}{2}$	$+\frac{1}{2}$
electron (e^-)	0.511 MeV	$\frac{1}{2}$	-1	0	$+1$	$+1$	0	0	0
positron (e^+)	0.511 MeV	$\frac{1}{2}$	$+1$	0	-1	-1	0	0	0
muon (μ^-)	106 MeV	$\frac{1}{2}$	-1	0	$+1$	0	$+1$	0	0
antimuon (μ^+)	106 MeV	$\frac{1}{2}$	$+1$	0	-1	0	-1	0	0
electron neutrino (ν_e)	< 2 eV[b,d]	$\frac{1}{2}$	0	0	$+1$	$+1$	0	0	0
electron antineutrino[a] ($\overline{\nu}_e$)	< 2 eV[b,d]	$\frac{1}{2}$	0	0	-1	-1	0	0	0
muon neutrino (ν_μ)	< 10 MeV[c,d]	$\frac{1}{2}$	0	0	$+1$	0	$+1$	0	0
muon antineutrino[a] ($\overline{\nu}_\mu$)	< 10 MeV[c,d]	$\frac{1}{2}$	0	0	-1	0	-1	0	0
charged pion ($\pi^\pm$)	140 MeV	0	± 1	0	0	0	0	1	± 1
neutral pion (π^0)	135 MeV	0	0	0	0	0	0	1	0

[a] It is not yet known experimentally whether neutrinos are different from antineutrinos.

[b] Measured in tritium beta decay for $\overline{\nu}_e$ and inferred for ν_e using CPT.

[c] Measured in $\pi^\pm$ decay.

[d] Cosmology gives model-dependent bounds on the sum of neutrino masses: $\lesssim 1$ eV. Neutrino-oscillation experiments indicate *differences* between neutrino masses are nonzero and much smaller than an eV.

For all scattering and decay process conservation of these quantities means that their sum over all particles in the initial state must agree with the sum over all particles in the final state.

2.2 Decays: general properties

A decay process is a reaction in which a single particle transmutes itself into two or more other particles, such as the reaction

$$P \to D_1 + D_2 + D_3 + \cdots + D_N \,, \tag{2.1}$$

[12] That is, it was thought that they were not exact in principle, but that in practice all non-conserving reactions were so small as to effectively never be observable.

in which the 'parent' particle, P, decays into N 'daughter' particles, D_1 through D_N. Such a decay is called an N-body decay because of the number of decay daughters present. (Normally energy and momentum conservation require $N \geq 2$.) A great many examples of decays of this type are observed since almost all known particles found in nature eventually decay, so it is the exception rather than the rule for a particle to be stable.

Sometimes the decay in question can be understood because the parent is built from smaller things and the decay represents either the decay of a constituent or the re-arrangement or escape of some of the constituents. Most nuclear decays (the source of ordinary radioactivity) turn out to be of this type:

- α *Decays:* correspond to the escape of a small He^4 nucleus (*i.e.* an α-particle) from a nucleus, and so always lower both Z and N by two (and so lower A by four): *e.g.* ${}^{238}_{92}\mathrm{U} \to {}^{234}_{90}\mathrm{Th} + \alpha$ is an example of a 2-body α-decay that is the start of the natural Uranium radioactive chain.

- β *Decays:* usually[13] correspond to the spontaneous decay of a neutron within the nucleus, whose occurrence is no surprise given that free neutrons are also seen to decay. The decay reaction for a free neutron is $n \to p^+ + e^- + \overline{\nu}_e$, (where p or p^+ denotes a proton and e^- denotes an electron while $\overline{\nu}_e$ is a particle called an 'electron anti-neutrino' — more about this later). Since the decay takes a neutron to a proton it always increases Z by one and leaves A unchanged, such as for the decay of radioactive Carbon: ${}^{14}_{6}\mathrm{C} \to {}^{14}_{7}\mathrm{N} + e^- + \overline{\nu}_e$ (which is the one used for carbon-14 radioactive dating). As we see later, in these decays the outgoing electron (and neutrino) are created at the instant of the decay and were not previously rattling around within the nucleus.

- γ *Decays:* correspond to the emission of a photon as the nuclear constituents fall from an excited energy level to one at lower energies (like the ground state), and so do not change Z or A at all. These are the nuclear analogs of the emission of light by atomic transitions wherein an excited electron jumps down to a lower energy level. The main difference is that nuclear γ transitions emit considerably more energy due to the larger energy differences between nuclear energy levels compared with atomic energy levels.

Neutron decay is an example where the bound state involved in a decay is not a nucleus but is instead something else. We shall see that neutrons (and more broadly all of the observed particles that take part in the nuclear force — *i.e.* protons, neutrons and many other unstable particles such as 'pions', $\pi^\pm$ or π^0) are built from still-smaller constituents called quarks, and some decays happen because of a decay of an underlying quark. For instance protons and neutrons are built from two types of quarks, u and d, with a proton being a 3-quark bound

[13] But not always: for some nuclei β decays can emit *positrons* — antiparticles to electrons — or sometimes they instead absorb an atomic electron into the nucleus.

state uud while a neutron is a udd state, and neutron decay arises because of the decay of an underlying quark: $d \to u + e^- + \overline{\nu}_e$.

> **Exercise 2.1:** Use conservation of electric charge, baryon number and lepton number, and the information that p is built as a uud bound state while n is a udd bound state, to infer the charge, baryon number and lepton number of the u and d quarks. What do these assignments mean for the charge, baryon number and lepton number assignments of the $\pi^{\pm}$ where π^+ is a $u\overline{d}$ combination and π^- is a $\overline{u}d$ combination?

Not all decays involve the rearrangements or decays of constituents, however, since particles that appear to be elementary are also known to decay. So far as we know the decay of a d quark is an example of this, as are other examples like the decay $\mu^- \to e^- + \overline{\nu}_e + \nu_\mu$ (where μ^- — called a 'muon' — is an elementary particle that is 200 times heavier than an electron, and ν_μ is a neutrino — called the 'muon neutrino' — that differs from the ν_e). It is the appearance of two types of neutrino here that make this reaction consistent with conservation of L_e and L_μ, since $L_e(\mu^-) = 0$ and $L_e(e^-) + L_e(\overline{\nu}_e) + L_e(\nu_\mu) = +1 - 1 + 0 = 0$ while $L_\mu(\mu^-) = +1$ and $L_\mu(e^-) + L_\mu(\overline{\nu}_e) + L_\mu(\nu_\mu) = 0 + 0 + 1 = +1$.

Neutron decays are also instructive for another reason. If neutrons are unstable the miracle is that any nuclei are stable at all, but (as we shall see) many are. Absolutely stable nuclei are stable because the increased Coulomb energy associated with the new proton's electric charge can make the prospective daughter nucleus heavier than the putative parent (and so the decay is forbidden by energy conservation). This observation teaches us (at least) two things: first, it shows that even for unstable nuclei the actual nuclear β-decay rate is not simply related to the decay rate of a free neutron. Nuclear decay rates range over many orders of magnitude in size precisely because their computation often requires detailed knowledge of the structure of the parent and daughter nucleus.

The stability of some nuclei — in spite of free-neutron decay — also shows that conservation laws can explain why some particles never decay. Energy conservation requires all daughter particles to be lighter than their parent and so the lightest particle carrying a conserved charge must be stable. So far as is known conservation laws are the reason why *all* of the stable elementary particles do not decay. For instance, electrons are absolutely stable because electric charge is conserved and there is no lighter particle carrying charge into which the electron can decay. Similarly, protons would be absolutely stable if baryon number, B, is conserved, but because we do not know whether B is exactly conserved there are doubts as to whether or not the proton is exactly stable. (Even if unstable its decay lifetime is known to be much longer than the age of the universe, so any failure in B conservation must be extremely small.)

2.2.1 Decay Kinematics

Conservation laws strongly restrict the properties of decays, and sometimes control whether or not a decay takes place at all. For instance for neutron decay, $n \to p + e^- + \overline{\nu}_e$, electric charge balances because a neutron is electrically neutral and so the initial total charge is $Q_i = Q(n) = 0$. For the decay products the proton and electron have opposite charge and the antineutrino is neutral, so $Q_f = Q(p) + Q(e^-) + Q(\overline{\nu}_e) = +e - e + 0 = 0$. Baryon number also balances because $B_i = B_f = +1$ with $B_i = B(n)$ carried by the decaying neutron while $B_f = B(p) + B(e^-) + B(\overline{\nu}_e) = +1 + 0 + 0$ is carried purely by the final proton. Lepton number is balanced because the initial neutron has $L_i = L(n) = 0$ while the final lepton number is $L_f = L(p) + L(e^-) + L(\overline{\nu}_e) = 0 + 1 - 1 = 0$. Lepton number would *not* be conserved if the antineutrino had instead been a neutrino, or if it carried $L = 0$.

Energy and momentum conservation similarly relate the initial and final states, with

$$E_P = \sum_{a=1}^{N} E_a \qquad \text{and} \qquad \mathbf{p}_P = \sum_{a=1}^{N} \mathbf{p}_a \,, \tag{2.2}$$

where E_a and $\mathbf{p}_a$ are the energy and momentum of particle D_a. Because each particle satisfies $E = \sqrt{\mathbf{p}^2 + m^2}$, where m is that particle's rest mass, and because the energy and the momentum of the initial particle can be chosen when setting up the experiment, there are a total of $3N$ unknowns (the components of the N final momenta) one would wish to determine. These unknowns are subject to the 4 constraints given in (2.2), and so in general we expect there to be $3N - 4$ free components of momentum that are not determined purely from energy-momentum conservation.

This counting is particularly simple in the rest-frame of the decaying particle, for which $\mathbf{p}_P = 0$ and so $E_P = m_P$. Consider first a two-body decay, $P \to D_1 + D_2$, for which $N = 2$ and so only $3N - 4 = 2$ of the 6 components of $\mathbf{p}_1$ and $\mathbf{p}_2$ are undetermined by energy-momentum conservation. In this case because $\mathbf{p}_P = 0$ momentum conservation requires the final momenta must sum to zero:

$$\mathbf{p}_1 + \mathbf{p}_2 = 0 \,, \tag{2.3}$$

so the two daughter particles emerge back-to-back in the decaying particle rest frame. This implies, in particular, that the magnitudes of their momenta are equal, and so therefore the energies of the two final particles in this frame must be related by

$$E_1^2 - m_1^2 = E_2^2 - m_2^2 \,. \tag{2.4}$$

But energy conservation also implies the energies in this frame must satisfy

$$E_1 + E_2 = m_P \,, \tag{2.5}$$

which only has solutions (for nonzero real momenta) if $m_1 + m_2 < m_P$. These two equations can be solved to determine the energies of both particles completely, giving:

$$E_1 = \frac{m_P^2 + m_1^2 - m_2^2}{2m_P} \qquad \text{and} \qquad E_2 = \frac{m_P^2 + m_2^2 - m_1^2}{2m_P}\,. \tag{2.6}$$

In particular, $E_1 \to E_2 \to \frac{1}{2}\, m_P$ in the limit $m_1 = m_2$ or when $m_1 \neq m_2$ but $m_1 + m_2 \ll m_P$.

> **Exercise 2.2:** Derive eq. (2.6) from eqs. (2.3) and (2.5) for two-body decay. Consider the decay $\pi^+ \to \mu^+ \nu_\mu$ and suppose muon neutrinos (ν_μ) emerge from π^+ decay and their energy (in the pion rest frame) is E_{ν_μ}. Suppose the mass, m_{ν_μ}, of the muon-neutrino (ν_μ) is inferred using measurements of E_{ν_μ} and the pion and muon masses, and that the result is consistent with vanishing m_{ν_μ}. What is the maximum error that can be allowed for each of m_π, m_μ and E_{ν_μ} in order for the 95% confidence limit on m_{ν_μ} to be 1 eV or smaller? (Assume the errors in each of these quantities is a Gaussian random variable and that all three are uncorrelated with one another.)
>
> [Bonus: consult the Particle Data Group webpage for the measured values of m_π and m_μ and their errors. Are these good enough to not prevent determining $m_{\nu_\mu} < 1$ eV at 95% confidence?]

Crucially: for two-body decays the energy of each of the decay products is completely determined (in any particular reference frame) by energy-momentum conservation. All that the details of the physics responsible for the decay can do is predict the likelihood for one of the particles to come out in a particular direction. (Even this is not possible if the initial parent particle is rotation invariant — *i.e* has no spin — since then all directions are equally likely.)

The same is not true when there are three or more particles in the final state. In this case momentum conservation in the decay rest frame implies

$$\mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 = 0\,, \tag{2.7}$$

which can be used to determine the momentum of one of the daughters in terms of the other two. But this does not fix the magnitude of each of the particle momenta separately, so energy can be shared between the other two particles consistently with overall energy conservation, which in the decay rest-frame states

$$E_1 + E_2 + E_3 = m_P\,. \tag{2.8}$$

Unlike for two-body decays, for three-body (or more-body) decays energy-momentum conservation is consistent with daughter particles emerging with a distribution of energies. For instance, in nuclear β-decay the electron is seen to emerge from the decay with a distribution of energies, as in Fig. 10, rather than a unique energy, and this is historically the way that the existence of the antineutrino was initially inferred.

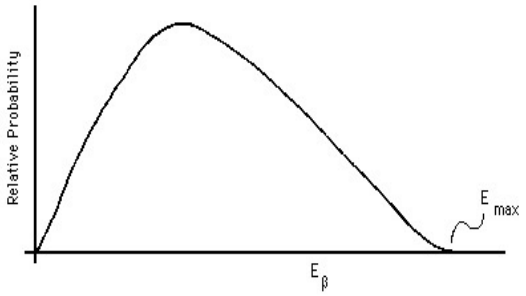


Figure 10. The distribution of electron energies obtained from the β decay of a nucleus. Because more than one energy is possible we know β decay cannot be 2-body and so must involve at least three particles in the final state. (Figure source: `http://www.ohio.edu/people/piccard/radnotes/alphabeta.html`).

2.2.2 Decay rates

The rest-frame decay probability per unit time (or decay rate), Γ, for a given particle is a characteristic of that particle as intrinsic to it as is its mass or spin. The value of Γ depends on the details of the interactions responsible for the decay, and because of this measurements of Γ can be informative about these interactions.

Because the decay of any one particle is a random event decay measurements involve statistical properties of a collection of identical decaying individuals, so we need the probability distribution, $\mathcal{P}(t)$, for a particle's continued survival at time t, given its existence at an initial time t_0. Now Γ is time-independent and $\Gamma\,\mathrm{d}t$ gives the probability for decay to occur in any given short time window, $\mathrm{d}t$, so the probability of there *not* being a decay in this interval is $1-\Gamma\,\mathrm{d}t$. Consequently the survival probability, $\mathcal{P}(t+\mathrm{d}t)=\mathcal{P}(t)+\mathrm{d}\mathcal{P}$, at time $t+\mathrm{d}t$ is given in terms of the survival probability, $\mathcal{P}(t)$, at t by

$$\mathcal{P}(t)+\mathrm{d}\mathcal{P}=\mathcal{P}(t)\,(1-\Gamma\,\mathrm{d}t) \qquad \text{and so} \qquad \mathrm{d}\mathcal{P}=-\Gamma\,\mathcal{P}(t)\,\mathrm{d}t\,, \tag{2.9}$$

where the factor $\mathcal{P}(t)$ after the first equality is the probability of surviving to time t, while the second factor is the probability of also surviving the next interval $\mathrm{d}t$.

The result $\mathrm{d}\mathcal{P}/\mathrm{d}t=-\Gamma\,\mathcal{P}$ integrates to give an exponential distribution

$$\mathcal{P}(t)=\Gamma\,e^{-\Gamma(t-t_0)}\,, \tag{2.10}$$

with an integration constant chosen to normalize the result on the interval $t_0\le t<\infty$. This shows that one way to measure Γ is by measuring the *mean life*, τ, of the decaying particles to survive (given their presence at the initial time t_0), defined by

$$\tau:=\int_{t_0}^{\infty}\mathrm{d}t\;(t-t_0)\,\mathcal{P}(t)=\frac{1}{\Gamma}\,. \tag{2.11}$$

Another way to get at the same quantity is by counting the population of undecayed particles in a sample as a function of time. For a collection of n_0 particles at time t_0 the number of surviving particles at a later time t is $n(t) \propto \mathcal{P}(t)$, and so

$$n(t) = n_0\, e^{-\Gamma(t-t_0)}\,. \tag{2.12}$$

Another measure of Γ is then the *half-life*, $\tau_{1/2}$, defined as the time taken for half of a given sample to decay (a result which doesn't depend on the size of the initial sample for exponential decays):

$$n(t - t_0 = \tau_{1/2}) = \frac{n_0}{2} \qquad \Longrightarrow \qquad \tau_{1/2} = \frac{\ln 2}{\Gamma} \simeq 0.693\,\tau\,. \tag{2.13}$$

The decay rate of a moving particle differs from the rest-frame rate because of time dilation. In a reference frame for which the decaying particle has speed $v = p/E$ the mean decay life is longer by the Lorentz transformation formula

$$\tau(E) = \gamma\,\tau = \frac{\tau}{\sqrt{1-v^2}} = \frac{E\tau}{m} \qquad \text{and so} \qquad \Gamma(E) = \frac{m\,\Gamma}{E}\,. \tag{2.14}$$

Time-dilation of decay lifetimes is a well-established experimental fact. For example, muons were initially discovered once radiation detectors were developed because we are constantly bombarded (several per square metre per second at the Earth's surface) by energetic muons coming down from the top of the atmosphere. Muons are produced there as byproducts of nuclear reactions when cosmic rays — *i.e.* mostly energetic protons — hitting the Earth from space collide with atomic nuclei in the upper atmosphere. But muons produced in the lab are seen to decay with lifetimes of about a microsecond, and at face value this causes a problem since even if moving at the speed of light a particle can move only 300 m in a microsecond. The problem is that the top of the atmosphere (where the muons are produced) is many kilometres up and so how can the muons live long enough to get down to the Earth's surface to be discovered? The resolution is time dilaton: although muons decay in microseconds in their rest frame, they are sufficiently energetic that they live long enough to an observer at rest on the Earth to survive the trip through the atmosphere.

2.2.3 Line widths

The above methods for measuring Γ are fine if the decay is slow enough. Nuclear decays are seen with an enormous range of half-lives — running from lifetimes in the billions of years down to lifetimes measured in small fractions of a second — so for many of these the above methods suffice. But many other decays are much faster: examples discussed in later sections can have $\Gamma \simeq 1$ GeV, and so $\tau = 1/\Gamma \simeq 1$ GeV$^{-1} \simeq 6 \times 10^{-25}$ s, and even moving at the speed of light a particle decaying this fast could only traverse about 0.2 fm – *i.e.* much less than the size of a nucleus. Decays this fast happen so quickly that the parent particle is not directly seen. How is Γ measured when this is so?

For decays this rapid we must zoom out a bit and consider the reaction that produced the decaying particle in the first place. For concreteness, suppose the decaying particle, P, is produced as an intermediate state in a 2-body reaction, $A+B \to C+D$, in which particles A and B are collided and are converted into particles C and D that are long-lived and so can be observed leaving the scene. We imagine that for at least some of these reactions this process occurs in two steps: $A+B \to P$ after which P decays through the process $P \to C+D$. (Such a two-step process is called a *resonant* channel.) We also allow that there may be several other reaction channels leading from $A+B$ to $C+D$ that do *not* require the creation of an intermediate P particle, which we call 'prompt' since they are usually faster (not needing to wait for P to decay). We imagine preparing the initial particles with specific energies, E_A and E_B, and measuring the final energies, E_C and E_D, (and sometimes also the directions) of the outgoing particles, and the goal is to identify whether the two-step decay reaction $A+B \to P \to C+D$ can be distinguished from any prompt reactions even if the decay is much too fast for the lifetime to be directly measured. And if they can be distinguished we wish to see how to infer the value of Γ for the decay.

Although the details are not important for what follows, there are a variety of specific reactions of this form that are of practical interest. One such is the reaction $e^+e^- \to e^+e^-$ or the reaction $e^+e^- \to \mu^+\mu^-$, which proceed both through the electromagnetic and the weak interactions. These reactions (and others) were studied in detail using electron-positron colliders, culminating with the Large Electron-Positron Collider (LEP, at CERN in Geneva) and the Stanford Linear Collider (SLC, in California) in the late 1980s and early 1990s. These experiments were the first to have sufficient energy to produce a Z boson (whose mass is 90 GeV) and for energies around 90 GeV the 'decay' version of the reaction takes place by first having $e^+e^- \to Z$, with the produced Z then decaying to the final state, $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$. A plot of measurements for a similar reaction ($e^+e^- \to \overline{q}q$, where q means any species of quark – a particle from which strongly-interacting particles turn out to be built) is given in Figure 11. The 'prompt' version of these reactions are understood as arising due to the exchange of photons and W bosons, but never through their direct production and decay. The resonant-decay channel only occurs[14] (for this choice of initial particles) for the Z.

Since the prompt and decay mechanisms are independent of one another but share the same initial and final state their amplitudes must be summed:

$$\mathcal{A}_{\rm tot}(AB \to CD) = \int_{-\infty}^{\infty} {\rm d}t_0 \, \Big[\mathcal{A}_{\rm pr}(t_0) + \mathcal{A}_{\rm dc}(t_0)\Big] \,, \tag{2.15}$$

where subscripts 'pr' and 'dc' denote the prompt and decay contributions, and there is an integration over the unknown time, t_0, when the reaction takes place. The reaction probabilities are then the squares of the amplitude, $\mathcal{P} = |\mathcal{A}|^2$ (as usual in quantum mechanics).

[14] It turns out there are other states besides the Z that can be produced and then decay, but these consist of quark-antiquark bound states and do not contribute much at these energies, for reasons that will become clear shortly.

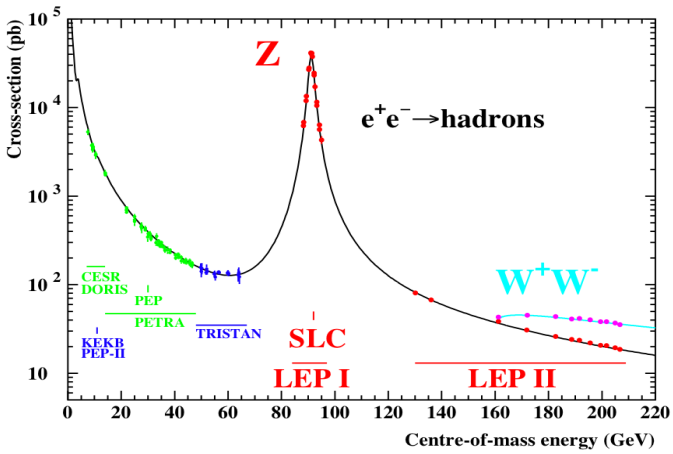


Figure 11. The Z boson resonance in the cross section for $e^+e^- \to$ hadrons (strongly interacting particles). The black line is the theoretical prediction of the Standard Model while the coloured points are measurements made at various accelerators. The large resonance peak occurs at the Z-boson mass, $M_Z \simeq 90$ GeV and its width characterizes the Z boson lifetime as described in the main text. Cross section is here measured in picobarns (pb), with a barn defined in eq. (1.36). (Figure source: `http://tlep.web.cern.ch/content/what-are-line-shape-parameters-resonance`).

For the decay contribution the amplitude comes with an additional sum over the time between production of P and its decay, since we do not know precisely when the decay occurs:

$$\mathcal{A}_{\rm dc}(t_0) = \int_{t_0}^{\infty} {\rm d}t\, \mathcal{A}(AB \to P; t_0)\, \mathcal{A}(P; t - t_0)\, \mathcal{A}(P \to CD; t)\,, \tag{2.16}$$

where $\mathcal{A}(AB \to P; t_0)$ is the amplitude for producing the intermediate state, P, at time t_0, $\mathcal{A}(P; t - t_0)$ is the amplitude for the P state to survive from t_0 until $t \geq t_0$ and $\mathcal{A}(P \to CD; t)$ is the amplitude for the decay $P \to C + D$ at time t. It is the delay $t - t_0$ caused by waiting for P to decay that allows us to distinguish the prompt from the decay reaction.

The important point is that the t- and t_0-dependence of these amplitudes is fairly simple to track. It arises in two ways: one is the contribution from the wave-function for each of the particles involved, $\psi_a \propto e^{-iE_a t}$, and since the amplitude is an inner product, $\mathcal{A} \sim \psi_f^* \psi_i$, this becomes $\psi_a^* \propto e^{+iE_a t}$ for any particle in the final state:

$$\begin{aligned} \mathcal{A}_{\rm pr}(t_0) &= \mathcal{A}_{AB|CD}\, e^{-i(E_A+E_B-E_C-E_D)t_0} \\ \mathcal{A}(AB \to P; t_0) &= \mathcal{A}_{AB|P}\, e^{-i(E_A+E_B-E_P)t_0} \\ \text{and} \qquad \mathcal{A}(P \to CD; t) &= \mathcal{A}_{P|CD}\, e^{-i(E_P-E_C-E_D)t}\,. \end{aligned} \tag{2.17}$$

The other source of t dependence is the exponential decay survival probability for P as it awaits its decay. That is,

$$\mathcal{A}(P; t - t_0) = i\mathcal{A}_P\, e^{-\Gamma(t-t_0)/2}\,, \tag{2.18}$$

where $\mathcal{A}_P$ is independent of time, and the factor of i is purely conventional. This last equation is required to allow the modulus of $\psi_P(t)$ to shrink exponentially, as it must since the survival probability is the square of the amplitude, $\mathcal{P}(P;t-t_0) \propto |\mathcal{A}(P;t-t_0)|^2 \propto e^{-\Gamma(t-t_0)}$, as required by the exponential decay law described above.

Consequently, the t- and t_0-dependence of the decay contribution to the integrand in (2.16) is

$$\mathcal{A}_{AB|P}\mathcal{A}_{P|CD}\mathcal{A}_P\, e^{-i(E_A+E_B-E_C-E_D)t_0}\, e^{-i[(E_P-i\Gamma/2)-E_C-E_D](t-t_0)}\,, \tag{2.19}$$

and this must be added to $\mathcal{A}_{\rm pr}$, whose t_0-dependence is given in (2.17). The t-integration is then elementary, and once this is done both prompt and decay terms share the common factor $e^{-i(E_A+E_B-E_C-E_D)t_0}$, whose integration with respect to t_0 gives the usual energy-conserving delta-function: $2\pi\,\delta(E_A+E_B-E_C-E_D)$. The total amplitude found by summing the prompt and decay processes then becomes

$$\mathcal{A}_{\rm tot}(AB\to CD) = 2\pi\left[\mathcal{A}_{AB|CD} + \frac{\mathcal{A}_{AB|P}\mathcal{A}_{P|CD}\mathcal{A}_P}{(E_P-E_{\rm tot})-i\Gamma/2}\right]\delta(E_A+E_B-E_C-E_D)\,, \tag{2.20}$$

where $E_{\rm tot}=E_A+E_B$ is the total energy available in the initial state.

Now comes the main point. The coefficients $\mathcal{A}_{AB|CD}$ and $\mathcal{A}_{AB|P}$ *etc* can depend on $E_{\rm tot}$, but this dependence is usually not particularly strong in the immediate vicinity of $E_{\rm tot}=E_P$. Consequently, if $\Gamma \ll E_P$ (a condition called the *narrow-resonance* condition), then it is the denominator in the second term of the square bracket that dominates the $E_{\rm tot}$-dependence of the rate very near $E_{\rm tot}=E_P$. Since its square has the form

$$\left|\frac{1}{(E_P-E_{\rm tot})-i\Gamma/2}\right|^2 = \frac{1}{(E_{\rm tot}-E_P)^2+(\Gamma/2)^2}\,, \tag{2.21}$$

this gives a large enhancement or resonance to the reaction rate, regarded as a function of $E_{\rm tot}$, occurring precisely at E_P. The $E_{\rm tot}$-dependence predicted in (2.21) is universal and is called the *Breit-Wigner* (or, in the relativistic context, *Lorentzian*) line-shape.

Whenever this characteristic line-shape is observed, the width Γ is easily determined from the width of the resonant peak, given for example by its full width at the place where the resonance has fallen off to half of its maximum value (*i.e.* the peak's 'full-width half-max'). For example, Fig. 11 plots the predicted and measured reaction rate for the reaction $e^+e^-\to$ hadrons (essentially quarks), which shows precisely this shape near the Z boson mass. Sometimes new particles are discovered by performing a 'bump hunt' that search for resonances whose presence could indicate the existence of something hitherto undiscovered. The most recent example of such a discovery is the Higgs boson, whose decay into two photons was discovered through the presence of an unexpected bump in the photon production rate at the Large Hadron Collider (LHC) at CERN. Figure 12 shows the Higgs bump due to the decay $H\to\gamma\gamma$.

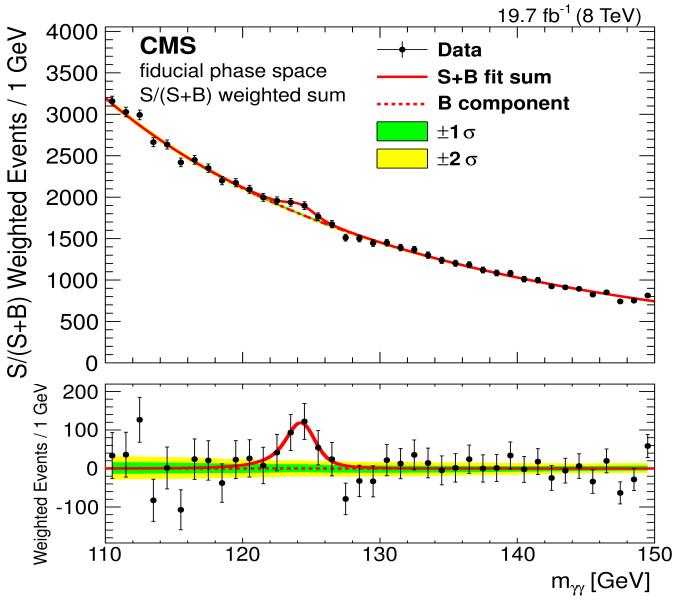


Figure 12. The Higgs boson resonance in the cross section for producing two photons in the CMS detector at the Large Hadron Collider. The top curve shows both prompt events and Higgs decays while the bottom panel is after the prompt events have been subtracted out. (Figure source: `http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-14-016/index.html`).

2.2.4 Differential decay rates

Decays into multiple daughters can sometimes give additional information, through the angular distribution of the decay products. This information is contained within the *differential decay rate*, which can be informative about the spin of the decaying particle and about the nature of the underlying interactions that are in play. The differential decay rate, $\mathrm{d}\Gamma/\mathrm{d}^3\mathbf{p}_1\cdots\mathrm{d}^3\mathbf{p}_N$, regarded as a function of $\mathbf{p}_1$ through $\mathbf{p}_N$, is the joint probability per unit time for a decay process with particle D_1 emitted within a small volume $\mathrm{d}^3\mathbf{p}_1$ of $\mathbf{p}_1$, particle D_2 emitted within a volume $\mathrm{d}^3\mathbf{p}_2$ about $\mathbf{p}_2$ and so on up to particle D_N having momentum within the volume $\mathrm{d}^3\mathbf{p}_N$ of $\mathbf{p}_N$. In terms of this the total decay rate discussed above is given by

$$\Gamma = \int \mathrm{d}^3\mathbf{p}_1\cdots\mathrm{d}^3\mathbf{p}_N \left(\frac{\mathrm{d}\Gamma}{\mathrm{d}^3\mathbf{p}_1\cdots\mathrm{d}^3\mathbf{p}_N}\right). \tag{2.22}$$

It is useful to define the differential decay rate in a way that is Lorentz invariant, since then it can be computed once and for all with the result useful for decays of particles in any reference frame. For the decay $P \to D_1(\mathbf{p}_1)+D_2(\mathbf{p}_2)+\cdots+D_N(\mathbf{p}_N)$ the invariant differential decay distribution, $\mathcal{M}(\mathbf{p}_P;\mathbf{p}_1,\cdots\mathbf{p}_N)$, is related to the above differential decay rate by

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}^3\mathbf{p}_1\cdots\mathrm{d}^3\mathbf{p}_N} = \frac{1}{2E_P}\left[\frac{\mathcal{M}(\mathbf{p}_P;\mathbf{p}_1,\cdots,\mathbf{p}_N)}{[(2\pi)^3 2E_1]\cdots[(2\pi)^3 2E_N]}\right](2\pi)^4\delta^4(p_P-p_1-p_2\cdots-p_N), \tag{2.23}$$

where the delta-function sets the sum of final 4-momenta, $p_1^\mu + \cdots + p_N^\mu$, equal to the initial 4-momentum, p_P^μ, so (in the centre-of-mass frame)

$$\delta^4(p_P - p_1 - \cdots p_N) := \delta(E_P - E_1 - E_2 \cdots - E_N)\, \delta^3(\mathbf{p}_1 + \cdots + \mathbf{p}_N)\,. \tag{2.24}$$

The Lorentz-invariance of $\mathcal{M}$ relies on the observation that Γ transforms as does m/E_P and so the rest must be invariant. Furthermore, the 4-dimensional delta-function is invariant since it imposes a relation amongst 4-momenta that all transform in the same way, and the measure $\mathrm{d}^3\mathbf{p}/E$ for each particle is also Lorentz-invariant, as can be seen by directly following through the transformations that take $\mathbf{p}$ and E to $\mathbf{p}'$ and E'.

> **Exercise 2.3:** Derive the transformation law for E and $\mathbf{p}$ (as a function of $\cosh\beta$ and $\sinh\beta$) for a boost along the z axis from the transformation law of the energy-momentum 4-vector, $(p')^\mu = \Lambda^\mu{}_\nu\, p^\nu$. By directly using these prove that $\mathrm{d}p_x$, $\mathrm{d}p_y$ and $\mathrm{d}p_z/E$ are invariant (from which we learn $\mathrm{d}^3\mathbf{p}/2E$ is also invariant, as claimed in the text).

Alternatively, the invariance of $\mathrm{d}^3\mathbf{p}/E$ can be seen by starting from the manifestly invariant starting point

$$\int \mathrm{d}^4p\, \delta(p_\mu p^\mu + m^2)\vartheta(p^0)(\cdots) = \int \mathrm{d}^3\mathbf{p}\, \mathrm{d}p^0\, \delta\left[-(p^0)^2 + E^2\right] \vartheta(p^0)(\cdots) = \int \frac{\mathrm{d}^3\mathbf{p}}{2E}\ (\cdots)|_{p^0=E}\,, \tag{2.25}$$

where $p_\mu p^\mu := \eta_{\mu\nu}\, p^\mu p^\nu$ so the delta-function imposes the condition $(p^0)^2 = E^2$ where $E := \sqrt{\mathbf{p}^2 + m^2}$ and the step-function, $\vartheta(x) = \{0 \text{ if } x < 0 \text{ and } 1 \text{ if } x > 0\}$, tells us to take the positive root when doing so. This condition on the sign of p^0 is also Lorentz-invariant because the delta-function tells us that p^μ is time-like (and so all observers agree on the sign of p^0).

Finally, the factor of $2E$ in the denominator of (2.25) arises from the change-of-variable formula for the Dirac delta-function, about which we pause to amplify because it is also useful later. Recall that $\delta(x-y)$ is defined to vanish for $x \neq y$ (and diverge for $x = y$) in such a way that

$$\int \mathrm{d}y\, \delta(x-y) f(y) = f(x) \tag{2.26}$$

for any integration region including $y = x$ and any sufficiently smooth function f. But the delta function in (2.25) instead comes in the form[15] $\int \mathrm{d}y\, \delta[g(x,y)]\, f(y)$ and so its evaluation requires a few extra steps:

$$\int \mathrm{d}y\, \delta[g(x,y)]\, f(y) = \int \frac{\mathrm{d}g}{|\partial g/\partial y|}\, \delta(g)\, f(y) = \left(\frac{f(y)}{|\partial g/\partial y|}\right)_{y=y(x)}\,, \tag{2.27}$$

[15] Explicitly, in the example of interest $y = p^0$, $x = E$ and $g(x,y) = x^2 - y^2$ so $g = 0$ implies $y = \pm x$ and $|\partial g/\partial y| = 2y$.

where the first equality changes the integration variable to agree with the argument of the δ-function (so as to use (2.26)), and $y = y(x)$ is the (assumed unique within the integration region) solution to $g(x, y) = 0$.

It is the invariant quantity $\mathcal{M}$ that we later relate to the square of a scattering amplitude once we try to compute the decay rate starting from an underlying theory of the interactions. Once this is done we will find

$$\mathcal{M} = \left\langle |\mathcal{A}|^2 \right\rangle , \tag{2.28}$$

where $\mathcal{A}$ is an invariant amplitude (often the matrix element of some interaction Hamiltonian, $\mathcal{A} = \langle f|H_{\rm int}|i\rangle$, between an initial state, $|i\rangle$, and a final state, $|f\rangle$) and $\langle \cdots \rangle$ denotes a sum over unmeasured quantum numbers (such as spin) in the final state, and an average over unmeasured quantum numbers in the initial state.

With these definitions, once the invariant quantity $\mathcal{M}$ is known, the total rate is computed using

$$\boxed{\mathrm{d}\Gamma(P \to F) = \frac{1}{2E_{_P}}\, \mathcal{M}\, (2\pi)^4 \delta^4(p_{_P} - p_{_F})\, \mathrm{d}\beta_{_F}} , \tag{2.29}$$

where $F = D_1 + \cdots + D_N$ here collectively denotes all of the final daughter particles, and so $p_{_F}$ is short-hand for the sum over final-state 4-momenta: $p^\mu_{_F} = p^\mu_1 + \cdots + p^\mu_{_N}$. Finally, the last factor denotes the combination

$$\mathrm{d}\beta_{_F} := \frac{\mathrm{d}^3\mathbf{p}_1}{(2\pi)^3 2E_1} \cdots \frac{\mathrm{d}^3\mathbf{p}_{_N}}{(2\pi)^3 2E_{_N}} . \tag{2.30}$$

The total rate is obtained by integrating over all possible final-state momenta, and because this volume of integration is called the reaction's *phase space*, the product in (2.30) is called the *Lorentz-invariant phase-space* (or LIPS) measure.

> **Exercise 2.4:** Evaluate the integrals over Lorentz-invariant phase space and show that for two-body decay the differential decay rate for emission of one of the daughters into an element of solid angle, $\mathrm{d}\Omega$, is given in the rest frame of the decaying particle by
>
> $$\frac{\mathrm{d}\Gamma}{\mathrm{d}\Omega}(A \to B + C) = \frac{\mathcal{M}\, p}{32\pi^2 m_{_A}^2} \qquad \text{(decay rest frame)} , \tag{2.31}$$
>
> where $p = \sqrt{E_{_C}^2 - m_{_C}^2} = \sqrt{E_{_D}^2 - m_{_D}^2}$ is the magnitude of the momentum of either of the daughter particles. Given the daughter energies are $E_{_B} = (m_{_A}^2 + m_{_B}^2 - m_{_C}^2)/2m_{_A}$ and $E_{_C} = (m_{_A}^2 + m_{_C}^2 - m_{_B}^2)/2m_{_A}$ show that this means
>
> $$p = \frac{\sqrt{[m_{_A}^2 - (m_{_B} + m_{_C})^2][m_{_A}^2 - (m_{_B} - m_{_C})^2]}}{2m_{_A}} . \tag{2.32}$$

Exercise 2.5: The charged pion, π^+, decays almost always into $\mu^+\nu_\mu$. It turns out the invariant matrix element for this decay is

$$\mathcal{M}\left(\pi^+ \to \mu^+\nu_\mu\right) = 2m_\pi p\Big(2G_{_F}|V_{ud}|\, m_\mu F_\pi\Big)^2 , \tag{2.33}$$

where p is the magnitude of the neutrino momentum in the decay rest frame, $m_\pi = 140$ MeV is the charged pion mass, $m_\mu = 105$ MeV is the muon mass and $G_{_F} = 1.166379 \times 10^{-5}$ GeV^{-2} is Fermi's constant and $|V_{ud}| = 0.974$ is called a Kobayashi-Maskawa matrix element. The quantity F_π is the pion decay constant, whose value is determined by the comparing this decay rate with the measured lifetime (once $G_{_F}$ is determined from μ^+ decay and $|V_{ud}|$ from nuclear β decay). Compute the total decay lifetime of the pion and show that it is given by

$$\Gamma(\pi^+ \to \mu^+\nu_\mu) = \frac{G_{_F}^2|V_{ud}|^2 m_\mu^2 m_\pi F_\pi^2}{4\pi}\left(1 - \frac{m_\mu^2}{m_\pi^2}\right)^2 . \tag{2.34}$$

Compare this to the measured mean life ($2.6033 \pm 0.0005 \times 10^{-8}$ sec), to see what the experimental value is for F_π. $\pi^+ \to e^+\nu_e$ can also occur, and does so with a rate obtained from the above by substituting $m_\mu \to m_e$ (where $m_e = 0.511$ MeV). What is the ratio $R_\pi = \Gamma(\pi^+ \to e^+\nu_e)/\Gamma(\pi^+ \to \mu^+\nu_\mu)$ numerically? Compare your answer with the experimental value for this ratio, which can be found at the Particle Data Group website (with information specifically about $\pi^\pm$ decays found here). Naively this ratio is something of a puzzle since electrons and muons participate in interactions with the same strength and the electron provides more phase space into which to decay, so one might have expected $R_\pi \gg 1$. The fact that this is not true tells us about the spin-dependence of the underlying weak interactions.

Exercise 2.6: The neutral pion, π^0, decays almost always into two photons. It turns out the invariant matrix element for this decay is

$$\mathcal{M}\left(\pi^0 \to \gamma\gamma\right) = 2\left[\frac{\alpha\, m_\pi^2}{2\pi F_\pi}\left(\frac{N_c}{3}\right)\right]^2 , \tag{2.35}$$

where $m_\pi = 135$ MeV is the neutral pion mass and $\alpha = 1/137$ is the fine-structure constant. The quantity $F_\pi = 92$ MeV is called the pion decay constant, and can be measured in the decay process $\pi^+ \to \mu^+\nu_\mu$. Finally, N_c is the number of colours carried by each quark inside the pion (more about which later). Compute the total decay rate of the pion and show it is

$$\Gamma\left(\pi^0 \to \gamma\gamma\right) = \frac{\alpha^2 m_\pi^3}{(4\pi)^3 F_\pi^2}\left(\frac{N_c}{3}\right)^2 . \tag{2.36}$$

(*Careful:* the two photons are completely indistinguishable. What is the proper solid angle through which one should integrate $\mathrm{d}\Gamma/\mathrm{d}\Omega$ if we are not to double-count?) Evaluate this and compare the result to the measured mean life ($8.52 \pm 0.18 \times 10^{-17}$ sec), to see what the experimental value is for the number of quark colours.

2.3 Scattering: general properties

The other major source of information about subatomic particles comes from studying collisions wherein the bringing together of several (in subatomic physics usually two) particles initiates a reaction of some sort, such as

$$A + B \to F_1 + F_2 + \cdots F_N \,, \tag{2.37}$$

which is a $2 \to N$ collision corresponding to having two particles collide with N particles leaving the reaction. *Elastic* collisions form the important special case of a $2 \to 2$ collision for which the final two particles are identical to the initial two: $A + B \to A + B$. All other collisions are called *inelastic*, because some of the initial kinetic energy has been converted into changing particle types. We next review the convenient ways to characterize the reaction rates for such collisions.

2.3.1 Cross sections and luminosity

Very rarely do experiments in subatomic physics prepare particles only one at a time for collisions, since normally a collection of particles are first accelerated to some energy in a high-energy beam before being brought to collide, either with another beam or with a stationary fixed target. Usually the more particles in the beam and target the more collisions there will be.

When particles collide there are two kinds of things that determine the reaction rate. Some of these are fairly mundane, like the number of particles involved (more particles means more potential reactions) and their speeds and other adjustable properties as they collide. Others are more fundamental, such as the interactions the particles experience. The goal of this section is to express the reaction rate for a collision in terms of an initial *luminosity* (which captures the mundane features specific to the particular way the particles were brought together) and an interaction *cross section* that contains the information about the interactions involved.

For these purposes consider a beam of particles containing n_B particles per unit area and moving with speed v relative to a target, and suppose the target is a large spherical object of radius R, with which an interaction occurs with probability p if the particles impinge on the target's surface (see Figure 13). The number of reactions occurring, $\mathrm{d}N_R$, in a small time window $\mathrm{d}t$ is then given by p times the number of particles arriving at this surface in time

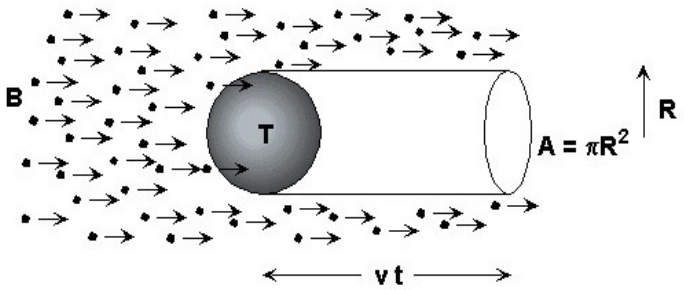


Figure 13. Schematic collision process for which a beam of small particles impinges onto a large spherical target of radius R. (Figure source: `http://www.jupiterscientific.org/sciinfo/crosssection.html`).

interval $\mathrm{d}t$, and so is $\mathrm{d}N_R = n_B(v\,\mathrm{d}t)(\pi R^2)p$. This makes the reaction rate

$$\frac{\mathrm{d}N_R}{\mathrm{d}t} = n_B v\,(\pi R^2)p = \mathcal{L}\,\sigma\,, \tag{2.38}$$

where $\mathcal{L} = n_B v$ gives the beam's *luminosity* — *i.e.* the number of particles per unit area per unit time delivered to the target — and $\sigma = p\,\pi R^2$ is called the interaction *cross section*, and has dimensions of area. Notice that in the simple scattering model used here σ simply is the area the target presents to the beam if the interaction probability per collision is $p = 1$, but is smaller otherwise. More generally (such as if target and beam interact at a distance through long-range forces, or once diffractive quantum wave behaviour is considered) it is possible also to have σ be larger than the target's cross-sectional area.

Instantaneous luminosity is a property of the accelerator that produces the beam, and a typical example from a modern accelerator might be of order $\mathcal{L} = 10^{34}$ cm^{-2} s^{-1}. *Integrated luminosity*, L, is another useful statistic that gives the total number of particles per unit area delivered on target over some time window (such as the lifetime of an experiment, say),

$$L(T) = \int_{t_0}^{t_0+T} \mathrm{d}t\;\mathcal{L}(t)\,, \tag{2.39}$$

which has units of inverse area. For instance, delivering the above luminosity for $T = 1$ year $\simeq 3\times10^7$ s gives an integrated luminosity $L \simeq 3\times10^{41}$ per square cm. Multiplying integrated luminosity times cross section, $N = \sigma L$, directly gives the total number of scattering events that occur over the given time window, T.

Of course we would be nuts to continue using CGS (or SI) units here, and for subatomic physics something closer to the dimensions of a nucleus makes a better reference unit. The conventional choice is the *barn* (or b), defined in (1.36) (and repeated here) as

$$1\text{ b} = 10^{-24}\text{ cm}^2 = 10^{-28}\text{ m}^2 = (10\text{ fm})^2\,, \tag{2.40}$$

together with the usual metric prefixes: mb, μb, nb, fb, pb and so on. These units are also useful for describing integrated luminosity, with $L = 1$ pb^{-1} corresponding to 10^{36}/cm^2. In

these units $L \simeq 3\times10^{41}$ cm^{-2} becomes $L \simeq 100$ inverse femtobarns, and so even cross sections as small as $\sigma \simeq 1$ fb would generate 100 events given this much integrated luminosity.

2.3.2 Invariant and differential cross section

One drawback of the previous section is that it is entirely phrased within the rest-frame of the target, and so the separation of the rate into a luminosity piece and a cross-section piece is not yet Lorentz invariant. This is a drawback because not all experiments are done with motionless targets (an example is a colliding beam experiment — like the LHC or LEP — which collide two beams into one another head on). This section aims in part to correct this drawback.

Furthermore, we are usually interested not just in the total cross section but also in the *differential* cross section, for which specific values of final-state momenta are specified for the outgoing particles. It is also useful to define this in a Lorentz-invariant way, making it easier to convert predictions to any particular frame of interest for a specific experiment.

The starting point for defining things covariantly is the reaction rate, Γ, and its differential counterpart

$$\Gamma(AB \to F_1 \cdots F_N) = \int \mathrm{d}^3\mathbf{p}_1 \cdots \mathrm{d}^3\mathbf{p}_N \left(\frac{\mathrm{d}\Gamma}{\mathrm{d}^3\mathbf{p}_1 \cdots \mathrm{d}^3\mathbf{p}_N}\right) . \tag{2.41}$$

For a two-particle initial state Γ transforms under Lorentz transformations like $1/(E_A E_B)$ — one way to see this is because the process $AB \to F_1 \cdots F_N$ could have been the independent decay of the initial particles A and B rather than a collision, and we have seen above that each decay rate separately transforms like $1/E$ for the particle decaying. Keeping in mind, as before, that the measure $\mathrm{d}^3\mathbf{p}/E$ is Lorentz-invariant suggests defining the *invariant scattering rate*, $\mathcal{M}(\mathbf{p}_A, \mathbf{p}_B; \mathbf{p}_1, \cdots, \mathbf{p}_N)$, by

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}^3\mathbf{p}_1 \cdots \mathrm{d}^3\mathbf{p}_N} = \frac{n_B}{2E_A 2E_B}\left[\frac{\mathcal{M}(\mathbf{p}_A, \mathbf{p}_B; \mathbf{p}_1, \cdots, \mathbf{p}_N)}{[(2\pi)^3 2E_1]\cdots[(2\pi)^3 2E_N]}\right] (2\pi)^4\delta^4(p_A + p_B - p_1 - p_2 \cdots - p_N) , \tag{2.42}$$

where, as before, n_B denotes the density of beam particles and the delta-function sets the sum of final 4-momenta, $p_1^\mu + \cdots + p_N^\mu$, equal to the initial 4-momentum, $p_A^\mu + p_B^\mu$. It is again $\mathcal{M} = \langle|\mathcal{A}^2|\rangle$ that is related to squares of scattering amplitudes computed using an underlying theory.

We can now use $\mathcal{M}$ to perform the split into luminosity and cross section in a way that makes the cross section also a Lorentz-invariant quantity. We do so by writing

$$\mathrm{d}\sigma = \frac{\mathrm{d}\Gamma}{\mathcal{F}} \tag{2.43}$$

as before, but now where $\mathcal{F}$ is chosen to: (*i*) agree with $\mathcal{L} = n_B v_{\rm rel}$ when the target (particle A, say) is at rest; and (*ii*) transform as does Γ to ensure $\mathrm{d}\sigma$ is Lorentz-invariant. Here $v_{\rm rel}$ — defined as the relative speed of the incident beam particles relative to the target — is itself a

Lorentz-invariant quantity, given in terms of the invariant dot product, $p_A \cdot p_B = \eta_{\mu\nu} p_A^\mu p_B^\nu \leq 0$, of initial 4-momenta by

$$v_{\rm rel} = \sqrt{1 - \frac{m_A^2 m_B^2}{(p_A \cdot p_B)^2}}\,. \tag{2.44}$$

Exercise 2.7: Prove the above relation for $v_{\rm rel}$ by evaluating the quantity $p_A \cdot p_B$ in terms of $v_{\rm rel}$ in the rest-frame of one of the particles, and then solving for $v_{\rm rel}$.

The solution to condition (ii) is $\mathcal{F} = n_B f/(2E_A 2E_B)$ where f is any Lorentz-invariant quantity (and the factors of 2 are conventional). Condition (i) then tells us

$$f = -4 v_{\rm rel}(p_A \cdot p_B) = 4\sqrt{(p_A \cdot p_B)^2 - m_A^2 m_B^2}\,, \tag{2.45}$$

because then $\mathcal{F} \to n_B v_{\rm rel}$ when $\mathbf{p}_A \to 0$.

There are two particularly useful frames of reference in $2 \to N$ scattering processes. One, usually called the *lab frame*, is the frame[16] in which one of the initial particles at rest. This is the frame within which our original discussion of luminosity and cross section was done. In the lab frame (rest-frame of B) and the c.o.m. frame f becomes

$$\begin{aligned} f = 4 m_B E_A v_{\rm rel} &= 4 m_B p_{A\,\rm lab} && \text{(lab frame)} \\ \text{and} \quad f = 4\sqrt{(E_A E_B + p_A^2)^2 - m_A^2 m_B^2} &= 4(E_A + E_B)_{\rm cm}\, p_{A\,\rm cm} && \text{(c.o.m. frame)}\,. \end{aligned} \tag{2.46}$$

The final expression for the invariant differential cross section then is

$$\boxed{\mathrm{d}\sigma(I \to F) = \frac{\mathcal{M}}{f}\,(2\pi)^4 \delta^4(p_I - p_F)\,\mathrm{d}\beta_F}\,, \tag{2.47}$$

where $I = A + B$ denotes the initial 2-body state and $p_I^\mu = p_A^\mu + p_B^\mu$ denotes the total initial 4-momentum, while (as before) $F = F_1 + \cdots F_N$ denotes all of the final-state particles and so $p_F^\mu = p_1^\mu + \cdots + p_N^\mu$. The Lorentz-invariant phase space measure, $\mathrm{d}\beta_F$, is given by (2.30).

2.3.3 $2 \to 2$ cross section

To make this more concrete let's work out $\mathrm{d}\sigma(AB \to CD)$ more explicitly for the special case of $2 \to 2$ scattering. In this case there are two particles in the final state, and so

$$\begin{aligned} \mathrm{d}\sigma(AB \to CD) &= \frac{\mathcal{M}}{f}\,(2\pi)^4 \delta^4(p_A + p_B - p_C - p_D)\,\mathrm{d}\beta_F \\ &= \frac{\mathcal{M}}{f}\,(2\pi)^4 \delta(E_A + E_B - E_C - E_D)\,\delta^3(\mathbf{p}_A + \mathbf{p}_B - \mathbf{p}_C - \mathbf{p}_D)\,\frac{\mathrm{d}^3\mathbf{p}_C}{(2\pi)^3 2E_C}\,\frac{\mathrm{d}^3\mathbf{p}_D}{(2\pi)^3 2E_D}\,. \end{aligned} \tag{2.48}$$

[16] The lab frame is indeed the reference frame of the laboratory in 'fixed-target' experiments in which a beam is collided with a stationary target. The lab frame need not be the rest frame of the physical laboratory, however, in collider experiments for which collisions occur between a pair of incident beams.

As stated earlier, the four conditions given by energy-momentum conservation have removed four of the six independent components of final momenta, and so for a 2-body final state we can take the two quantities undetermined by conservation laws to be the angles specifying the direction of the momentum of one of the outgoing particles: particle C, say.

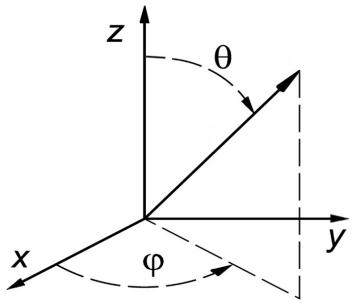


Figure 14. Definition of the angles θ and ϕ for the outgoing momentum $\mathbf{p}_C$. (Figure source: `https://commons.wikimedia.org/wiki/Spherical_polar_coordinates.png`).

We now perform the integral over the delta functions explicitly. We start with the integral over one of the two final momenta, say $\mathbf{p}_D$, whose integration with the momentum-conserving delta function amounts to everywhere replacing $\mathbf{p}_D$ with $\mathbf{p}_A + \mathbf{p}_B - \mathbf{p}_C$. Next write out the $\mathrm{d}^3\mathbf{p}_C$ integral in polar coordinates using $\mathrm{d}^3\mathbf{p} = \mathrm{d}p_x \mathrm{d}p_y \mathrm{d}p_z = p^2 \mathrm{d}p\, \mathrm{d}\Omega$ where $p = |\mathbf{p}|$ and the differential element of solid angle is $\mathrm{d}\Omega = \sin\theta\, \mathrm{d}\theta \mathrm{d}\phi$ with (θ, ϕ) giving the direction of $\mathbf{p}$, as in Figure 14. The energy-conserving delta function then allows us also to perform the integral over $|\mathbf{p}_C|$ as well, leaving only the angular integrals undone. When doing the $|\mathbf{p}_C|$ integral care must be used to properly use (2.27) for changing variables with the δ-function, leading to the result

$$\begin{aligned}
(2\pi)^4\, \delta^4(p_I - p_F)\, \mathrm{d}\beta_F &= (2\pi)^4\, \delta^4(p_A + p_B - p_C - p_D)\, \frac{\mathrm{d}^3\mathbf{p}_C}{(2\pi)^3 2E_C}\, \frac{\mathrm{d}^3\mathbf{p}_D}{(2\pi)^3 2E_D} \\
&= 2\pi\, \delta(E_A + E_B - E_C - E_D)\, \frac{\mathrm{d}^3\mathbf{p}_C}{(2\pi)^3 4E_C E_D}\bigg|_{\mathbf{p}_D = \mathbf{p}_A + \mathbf{p}_B - \mathbf{p}_C} \\
&= \frac{p_C^2\, \mathrm{d}\Omega_C}{(2\pi)^2 4E_C E_D |\mathrm{d}(E_C + E_D)/\mathrm{d}p_C|}\bigg|_{\mathbf{p}_D = \mathbf{p}_A + \mathbf{p}_B - \mathbf{p}_C,\, E_C = E_A + E_B - E_D} \\
&= \frac{p_C^3\, \mathrm{d}\Omega_C}{(4\pi)^2 |(E_D \mathbf{p}_C - E_C(\mathbf{p}_A + \mathbf{p}_B)) \cdot \mathbf{p}_C|}\bigg|_{\mathbf{p}_D = \mathbf{p}_A + \mathbf{p}_B - \mathbf{p}_C,\, E_C = E_A + E_B - E_D}.
\end{aligned} \tag{2.49}$$

and so the differential cross section for $2 \to 2$ scattering is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(AB \to CD) = \left[\frac{\mathcal{M}\, p_C^3}{(4\pi)^2 f |(E_D \mathbf{p}_C - E_C(\mathbf{p}_A + \mathbf{p}_B)) \cdot \mathbf{p}_C|}\right]_{\mathbf{p}_D = \mathbf{p}_A + \mathbf{p}_B - \mathbf{p}_C,\, E_C = E_A + E_B - E_D}, \tag{2.50}$$

where the right-hand side is to be regarded as a function of the direction, (θ, ϕ), of the outgoing momentum $\mathbf{p}_C$. The total cross section, σ, is then obtained by integrating this result over all possible such directions.

2.3.4 Lab and centre-of-mass frames

In the lab frame we can take $\mathbf{p}_B = 0$ and $E_B = m_B$, and so

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(AB \to CD) &= \left[\frac{\mathcal{M}\, p_C^3}{(4\pi)^2 f |(E_D \mathbf{p}_C - E_C \mathbf{p}_A)\cdot \mathbf{p}_C|}\right]_{\mathbf{p}_D = \mathbf{p}_A - \mathbf{p}_C,\, E_C = E_A + m_B - E_D} \qquad \text{(lab frame)} \\ &= \left[\frac{\mathcal{M}\, p_C^2}{(8\pi)^2 m_B p_A |E_D p_C - E_C p_A \cos\theta|}\right]_{\mathbf{p}_D = \mathbf{p}_A - \mathbf{p}_C,\, E_C = E_A + m_B - E_D}, \qquad (2.51)\end{aligned}$$

which uses $f = 4 v_{\mathrm{rel}} m_B E_A = 4 m_B p_A$ in the lab frame. In the special case where the incident particle (and its scattered partner) is massless, $E_A = p_A$ and $E_C = p_C$ this becomes

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(AB \to CD) = \left[\frac{\mathcal{M}\, E_C}{(8\pi)^2 m_B E_A |E_D - E_A \cos\theta|}\right]_{\mathbf{p}_D = \mathbf{p}_A - \mathbf{p}_C,\, E_C = E_A + m_B - E_D}, \qquad (2.52)$$

For most purposes a much more convenient frame is the *centre-of-mass frame* (or c.o.m. frame), defined by the condition that $\mathbf{p}_I := \mathbf{p}_A + \mathbf{p}_B = 0$. This frame is particularly simple both because it implies $|\mathbf{p}_A| = |\mathbf{p}_B|$ (and so also $E_A^2 - m_A^2 = E_B^2 - m_B^2$), and also because with momentum conservation it also implies $\mathbf{p}_C + \mathbf{p}_D = 0$ (and so $E_C^2 - m_C^2 = E_D^2 - m_D^2$).

As an example of how things often simplify in the c.o.m. frame, consider expression (2.50). In this frame we have $(E_D \mathbf{p}_C - E_C \mathbf{p}_D)\cdot \mathbf{p}_C = (E_D + E_C)\mathbf{p}_C \cdot \mathbf{p}_C = E_I\, p_C^2$, where the initial total energy is $E_I = E_A + E_B = E_C + E_D$. As a result (2.50) simplifies to become

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(AB \to CD) &= \left[\frac{\mathcal{M}\, p_C}{(4\pi)^2 f (E_A + E_B)}\right]_{\mathbf{p}_D = -\mathbf{p}_C,\, E_C = E_A + E_B - E_D} \qquad (2.53) \\ &= \left[\frac{\mathcal{M}\, p_C}{(8\pi)^2 p_A (E_A + E_B)^2}\right]_{\mathbf{p}_D = -\mathbf{p}_C,\, E_C = E_A + E_B - E_D} \qquad \text{(c.o.m. frame)}\,,\end{aligned}$$

Because $\mathbf{p}_A = -\mathbf{p}_B$ in the c.o.m. frame the initial momenta are parallel to one another, and so we can choose the direction they define to be the z-axis. In this case the angles (θ, ϕ) describe the direction of the line defined by the parallel final-state momenta relative to this initial direction. With this choice shifting ϕ corresponds to rotating the collision about the axis defined by the initial beam. It is often true that the physics is invariant under such a rotation, and when this is so the cross section is independent of ϕ and so depends nontrivially only on θ. In this case the angular integral over ϕ amounts to multiplication of the result by 2π, leaving

$$\begin{aligned}\frac{\mathrm{d}\sigma}{\sin\theta_C \mathrm{d}\theta_C}(AB \to CD) &= 2\pi \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)(AB \to CD) \qquad \text{(axially symmetric)} \qquad (2.54) \\ &= \left[\frac{\mathcal{M}\, p_C}{32\pi\, p_A (E_A + E_B)^2}\right]_{\mathbf{p}_D = -\mathbf{p}_C,\, E_C = E_A + E_B - E_D} \qquad \text{(c.o.m. frame)}\,.\end{aligned}$$

When this is true then there is only one independent final-state variable, θ, on which cross sections can nontrivially depend (in addition to their dependence on the choice of total initial energy, $E_{\rm cm} = E_A + E_B$).

2.3.5 $2 \to 2$ relativistic variables

Although formulae like (2.50) and (2.53) have the virtue of explicitness, they obscure Lorentz invariance and so make it more cumbersome to relate observables in different reference frames. For this purpose an alternative set of explicitly Lorentz-invariant variables, called *Mandelstam variables*, are often used instead of θ and ϕ to describe $2 \to 2$ scattering.

The Mandelstam variables are built directly in terms of the 4-momenta: p_A^μ, p_B^μ, p_C^μ and p_D^μ, and start with the observation that any Lorentz-invariant function of momenta (such as $\mathcal{M}$, for instance) can always be written as a function of the invariant inner products of these four 4-vectors: *e.g.* $p_A \cdot p_B = \eta_{\mu\nu} p_A^\mu p_B^\nu$. Because the inner product of a 4-momentum with itself is always the corresponding particle mass, $p_A \cdot p_A = -m_A^2$ and so on, they are constants and the only possible independent kinematic variables must be

$$p_A \cdot p_B\,, \quad p_A \cdot p_C\,, \quad p_A \cdot p_D\,, \quad p_B \cdot p_C\,, \quad p_B \cdot p_D \quad \text{and} \quad p_C \cdot p_D\,. \tag{2.55}$$

Even these are not all independent because, for example, 4-momentum conservation implies we can always eliminate p_D^μ using $p_A^\mu + p_B^\mu = p_C^\mu + p_D^\mu$, leaving three possible independent combinations like $p_A \cdot p_B$, $p_A \cdot p_C$ and $p_B \cdot p_C$. The conventional way to group these three quantities is into the Mandelstam variables s, t and u defined by

$$\begin{aligned} s &:= -(p_A + p_B) \cdot (p_A + p_B) = -2p_A \cdot p_B + m_A^2 + m_B^2\,, \\ t &:= -(p_A - p_C) \cdot (p_A - p_C) = +2p_A \cdot p_C + m_A^2 + m_C^2\,, \\ \text{and} \quad u &:= -(p_A - p_D) \cdot (p_A - p_D) = +2p_A \cdot p_D + m_A^2 + m_D^2\,. \end{aligned} \tag{2.56}$$

But we know that energy-momentum conservation and axial symmetry should only allow us two independent variables, the total initial energy and scattering angle in the c.o.m., for example. So we expect that even these three quantities, s, t and u, cannot really be independent. This expectation is right, and the relationship between them can be seen by summing the definitions to find $s + t + u = 2p_A \cdot (-p_B + p_C + p_D) + 3m_A^2 + m_B^2 + m_C^2 + m_D^2$, and then using 4-momentum conservation and $2p_A \cdot p_A = -2m_A^2$ to find

$$s + t + u = m_A^2 + m_B^2 + m_C^2 + m_D^2\,, \tag{2.57}$$

which allows us to eliminate u, say, in terms of s and t.

Evaluating the definitions in the c.o.m. frame shows how s and t are related to the two basic kinematic variables, $E_{\rm cm}$ and θ. Because $\mathbf{p}_A + \mathbf{p}_B = 0$ in this frame, the 4-vector $p_A^\mu + p_B^\mu$ points purely in the time direction, and so

$$s = (E_A + E_B)^2 = E_{\rm cm}^2 \qquad \text{(c.o.m. frame)}\,. \tag{2.58}$$

The energy of each particle separately is then detemined by the conditions that $p_A = p_B$ and $p_C = p_D$ while $E_A + E_B = E_C + E_D = E_{\rm cm}$. Because these conditions are essentially those that led to (2.6) they have the same solutions:

$$E_A = \frac{E_{\rm cm}^2 + m_A^2 - m_B^2}{2E_{\rm cm}} \qquad \text{and} \qquad E_B = \frac{E_{\rm cm}^2 + m_B^2 - m_A^2}{2E_{\rm cm}} \qquad \text{(c.o.m. frame)}\,, \tag{2.59}$$

and the identical equations with $(A, B) \to (C, D)$. Alternatively, evaluating $s = -2p_A \cdot p_B + m_A^2 + m_B^2$ in the lab frame (for which $\mathbf{p}_A = 0$) instead gives

$$s = 2m_A E_B + m_A^2 + m_B^2 \qquad \text{(lab frame)}\,. \tag{2.60}$$

Clearly $E_B \geq m_B$ implies $s \geq (m_A + m_B)^2$.

On the other hand evaluating $t = 2p_A \cdot p_C + m_A^2 + m_C^2$ in any frame relates it to the scattering angle, θ_C, between the direction of the outgoing particle C relative to the direction of the incoming particle A:

$$t = -2E_A E_C + 2\mathbf{p}_A \cdot \mathbf{p}_C + m_A^2 + m_C^2 = -2E_A E_C + 2p_A p_C \cos\theta_C + m_A^2 + m_C^2\,, \tag{2.61}$$

and this is particularly simple to use in the c.o.m. frame due to the explicit expressions like (2.59) for the energies (together with $p = \sqrt{E^2 - m^2}$ for each particle). Notice that the relation between t and θ is particularly simple in the ultra-relativistic limit, for which $E \simeq p \gg m$ for all particles. Then (2.61) degenerates to

$$\begin{aligned} t &\simeq -2E_A E_C(1 - \cos\theta_C) = -4E_A E_C \sin^2\frac{\theta_C}{2} \qquad \text{(ultra-relativistic)} \\ \text{and} \quad t &\simeq -\frac{E_{\rm cm}^2}{2}(1 - \cos\theta_C) = -E_{\rm cm}^2 \sin^2\frac{\theta_C}{2} \qquad \text{(ultra-relativistic c.o.m.)}\,. \end{aligned} \tag{2.62}$$

These last formulae show that $-s \leq t \leq 0$, and so is strictly non-positive, at least in the ultra-relativistic limit. They also show that for generic scattering angles $t \sim E_{\rm cm}^2$ is generically similar in size (but opposite in sign) to s, but also that this is not true for small enough scattering angles (*i.e.* $-t \simeq E_A E_C \theta_C^2 \ll E_{\rm cm}^2$ for $\theta_C \ll 1$).

We shall find that because $\mathcal{M}$ is Lorentz invariant it can be compactly written as a function of the Mandelstam variables, $\mathcal{M} = \mathcal{M}(s, t)$. The same is true of f, since (2.45) can be re-expressed as

$$f(s) = 4p_{A\,\rm cm}\sqrt{s} = 2\sqrt{[s - (m_A + m_B)^2]\,[s - (m_A - m_B)^2]}\,. \tag{2.63}$$

For this reason it is useful also to trade $\sin\theta_C\,{\rm d}\theta_C$ for ${\rm d}t$ and compactly express the differential cross section entirely in a manifestly Lorentz invariant way.

> **Exercise 2.8:** Use the definitions of s, t and u in the c.o.m. frame to derive the following useful expression for the differential Lorentz-invariant phase space

volume appearing in the cross section:

$$\begin{aligned} \mathrm{d}\chi &:= (2\pi)^4\,\delta^4(p_A + p_B - p_C - p_D)\,\frac{\mathrm{d}^3\mathbf{p}_C}{(2\pi)^3 2E_C}\,\frac{\mathrm{d}^3\mathbf{p}_D}{(2\pi)^3 2E_D} \\ &= -\delta(s+t+u-m_A^2-m_B^2-m_C^2-m_D^2)\,\frac{\mathrm{d}t\,\mathrm{d}u}{8\pi\xi(s)}\,, \end{aligned} \tag{2.64}$$

where

$$\xi(s) = 2\,p_{A\,\mathrm{cm}}\sqrt{s} = \sqrt{(s-m_A^2-m_B^2)^2 - 4m_A^2 m_B^2} = \frac{1}{2}\,f(s)\,, \tag{2.65}$$

and so $\xi(s) \to s$ in the ultra-relativistic limit, where $s \gg m_A^2,\, m_B^2$.

The results of the exercise allow the following manifestly invariant form for the differential cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t\,\mathrm{d}u}(AB\to CD) = -\frac{\mathcal{M}}{8\pi\xi(s)f(s)}\;\delta\left(s+t+u-m_A^2-m_B^2-m_C^2-m_D^2\right)\,, \tag{2.66}$$

or, using the δ-function to integrate over u and using the expressions for ξ and f,

$$\boxed{-\frac{\mathrm{d}\sigma}{\mathrm{d}t}(AB\to CD) = \frac{\mathcal{M}}{64\pi s\,p_{A\,\mathrm{cm}}^2} = \frac{\mathcal{M}}{16\pi\,[s-(m_A+m_B)^2]\,[s-(m_A-m_B)^2]}}\,. \tag{2.67}$$

Exercise 2.9: In Quantum Electrodynamics (QED) the process $e^+e^-\to\mu^+\mu^-$ takes place with an invariant amplitude $\mathcal{M}$ given by

$$\mathcal{M}(e^+e^-\to\mu^+\mu^-) = \frac{32\pi^2\alpha^2}{s^2}\Big(u^2+t^2\Big)\,, \tag{2.68}$$

in the ultra-relativistic regime where s, t, and u are much larger than the electron and muon masses. (This regime is a very good approximation for most applications to modern accelerators.) Here $\alpha = e^2/4\pi\hbar c \simeq 1/137$ is the dimensionless *fine-structure constant*. Compute $\mathrm{d}\sigma/\mathrm{d}u\mathrm{d}t$ as a function of s, t and u. Use your result to compute $\mathrm{d}\sigma/\mathrm{d}\Omega$ in the c.o.m. frame. Is the result you find isotropic? Integrate the differential cross section and show that the total cross section is $\sigma_{\mathrm{tot}} = 4\pi\alpha^2/(3s)$. What is σ_{tot} in nanobarns for $E_{\mathrm{cm}} = 10$ GeV?

Exercise 2.10: The process $e^-\mu^-\to e^-\mu^-$ in QED is characterized by the following invariant amplitude

$$\mathcal{M}(e^-\mu^-\to e^-\mu^-) = \frac{32\pi^2\alpha^2}{t^2}\Big(u^2+s^2\Big)\,, \tag{2.69}$$

in the ultra-relativistic regime where s, t, and u are much larger than the electron and muon masses. As in the previous problem $\alpha = e^2/4\pi\hbar c$ the dimensionless fine-structure constant. (Notice that $\mathcal{M}$ for this problem differs from the corresponding quantity in the previous problem only by the interchange $t \leftrightarrow s$, a special case of a general result known as 'crossing symmetry'.) Compute $\mathrm{d}\sigma/\mathrm{d}u\mathrm{d}t$ as a function of s, t and u. Use your result to compute $\mathrm{d}\sigma/\mathrm{d}\Omega$ in the c.o.m. frame. Compare your result with the Rutherford scattering cross section — see for instance eq. (3.22). Does your result agree on the size and angular dependence? If not is there a limit for which it does agree?

Exercise 2.11: In the Standard Model the invariant rate for the process $e^+e^- \to \mu^+\mu^-$ is given near the Z resonance (*i.e.* E_{cm} around 90 GeV) by

$$\mathcal{M}(e^+e^- \to \mu^+\mu^-) = \frac{(4\pi\alpha_z)^2}{|s - M^2 - iM\Gamma|^2}\left[\left(g_L^4 + g_R^4\right)u^2 + 2g_L^2 g_R^2 t^2\right], \qquad (2.70)$$

where we are in the ultra-relativistic regime where we drop electron and muon masses compared with s, t, and u. In this expression $\alpha_z = \alpha/s_w^2 c_w^2$ with $s_w = \sin\theta_W$ and $c_w = \cos\theta_W$ a parameter of the theory. M and Γ denote the mass and total decay rate of the Z boson. Finally, the couplings g_L and g_R are the left- and right-handed couplings of the electron and muon to the Z, given by $g_L = -\frac{1}{2} + s_w^2$ and $g_R = s_w^2$.

Compute $\mathrm{d}\sigma/\mathrm{d}u\mathrm{d}t$ as a function of s, t and u. Use your result to compute $\mathrm{d}\sigma/\mathrm{d}\Omega$ in the c.o.m. frame. Does your result have the Breit-Wigner factor — see for instance eq. (2.21). Use $M \simeq 90$ GeV, $\Gamma \simeq 2.4$ GeV and $s_w^2 = 0.23$ to compute the total cross section, σ_{tot}, and evaluate the result in nanobarns.

Exercise 2.12: The process $e^+e^- \to \mu^+\mu^-$ proceeds both through both the 'prompt' process described in Exercise 2.9 and the 'resonant' process in Exercise 2.11. Evaluate the total cross section for each of these processes separately at the energy $E = M = 90$ GeV. If an accelerator delivers a luminosity $\mathcal{L} = 10^{32}$ cm^{-2} sec^{-1} what is the event rate expected at this energy for each of these two processes? Assuming the rate for each of these processes can be calculated separately from one another,[17] calculate how long the experiment must run in order to distinguish the resonant process from a 3-standard-deviation (or 3-σ) fluctuation in the prompt process. (The result of Exercise 1.9 is useful when doing so.) What integrated luminosity is required to distinguish the resonant process from a 5-σ fluctuation of the prompt process?

[17] In reality the total rate is given by summing amplitudes rather than their rates, but this difference turns out not to matter for this question because these processes do not interfere right at resonance.

3 Calculational tools II

We now turn to calculating a few cross sections from known interactions in order to see what measurements of cross sections can tell us about the underlying interactions at play.

3.1 Classical two-body scattering

We start in this section with several examples calculated using classical Newtonian physics. Besides being instructive in their own right and providing a baseline against which to compare later quantum calculations, they also include examples of practical historical interest such as the Coulomb scattering cross section used by Rutherford.

3.1.1 Reduction to a one-body problem

Consider two particles that move on classical trajectories, $\mathbf{r}_A(t)$ and $\mathbf{r}_B(t)$, and mutually interact through a conservative force described by a central potential $V = V(r)$, where $r = |\mathbf{r}_A - \mathbf{r}_B|$ is the distance between the two particles. We know each satisfies Newton's 2nd Law, and so denoting time derivatives by over-dots, $\mathbf{v} = \dot{\mathbf{r}}$ and $\mathbf{a} = \dot{\mathbf{v}} = \ddot{\mathbf{r}}$, we have

$$m_A\,\ddot{\mathbf{r}}_A = \mathbf{F}_{AB} \qquad \text{and} \qquad m_B\,\ddot{\mathbf{r}}_B = \mathbf{F}_{BA}\,, \tag{3.1}$$

where Newton's 3rd Law implies $\mathbf{F}_{AB} = -\mathbf{F}_{BA}$.

The sum of these equations tells us the centre-of-mass moves in a straight line, $\ddot{\mathbf{R}} = 0$, where $(m_A + m_B)\mathbf{R} = m_A\mathbf{r}_A + m_B\mathbf{r}_B$, so we can simplify the description of the relative motion of the two particles by referring it to the c.o.m. frame, writing

$$\mathbf{r}_A = \mathbf{R} + \frac{m_B\,\mathbf{r}}{m_A + m_B} \qquad \text{and} \qquad \mathbf{r}_B = \mathbf{R} - \frac{m_A\,\mathbf{r}}{m_A + m_B}\,, \tag{3.2}$$

where $\mathbf{r} := \mathbf{r}_A - \mathbf{r}_B$. Eq. (3.1) then implies that the equation of motion for $\mathbf{r}(t)$ makes no reference to $\mathbf{R}$ and corresponds to motion of a single particle within a central potential, $U(r)$:

$$m\,\ddot{\mathbf{r}} = \mathbf{F}_{AB} = -\nabla U(r) = -\left(\frac{\mathrm{d}U}{\mathrm{d}r}\right)\mathbf{e}_r\,, \tag{3.3}$$

with $\mathbf{e}_r = \mathbf{r}/r$ the unit vector pointing in the radial direction, and the *reduced mass*, m, defined by $m^{-1} = m_A^{-1} + m_B^{-1}$.

> **Exercise 3.1:** Consider elastic $2 \to 2$ for which the particle type does not change, such as $\mu^+ + e^- \to \mu^+ + e^-$ or $\mu^+ + N \to \mu^+ + N$ for a nucleus N. (Both of these are special cases of the process $A + B \to C + D$ with $m_C = m_A$ and $m_D = m_B$). Define the lab frame as the frame in which particle B (the electron or the nucleus) is initially not moving, $\mathbf{p}_B = 0$, and choose the initial momentum for the (anti)muon to be along the positive z-axis: $\mathbf{p}_A = p\,\mathbf{e}_z$, with initial lab-frame energy $E_\mu^{\rm lab} := E_A^{\rm lab} = \sqrt{p^2 + m_\mu^2}$ (where $m_\mu \simeq 105$ MeV is the muon mass).

Use the invariance of the Mandelstam variable s to derive a formula for the total centre-of-mass energy (*i.e.* $E_{\rm cm} = E_A^{\rm cm} + E_B^{\rm cm}$ in the c.o.m. frame) as a function of $E_\mu^{\rm lab}$. Derive eqs. (2.59) and thereby compute the final energies $E_\mu^{\rm cm\prime} = E_C^{\rm cm}$ and $E_{N,e}^{\rm cm\,\prime} = E_D^{\rm cm}$ in the centre-of-mass frame as a function of $E_\mu^{\rm lab}$. U

Prove that the Mandelstam variable u defined in (2.56) satisfies $u = -(p_B - p_C)^2$. Use the Lorentz-invariance of u to compute the lab-frame energy $E_C^{\rm lab}$ as a function of $E_{\rm cm}$ and the c.o.m.-frame scattering angle θ (defined by $\mathbf{p}_A \cdot \mathbf{p}_C = p_A p_C \cos\theta$), and thereby compute the final lab-frame antimuon energy, $E_\mu^{\rm lab\prime}$, as a function of its initial energy, $E_\mu^{\rm lab}$, and the c.o.m. scattering angle.

Exercise 3.2: Consider again the elastic $2 \to 2$ scattering ($\mu^+ + e^- \to \mu^+ + e^-$ or $\mu^+ + N \to \mu^+ + N$ for a nucleus N) considered in the previous exercise. (Both of these are special cases of the process $A + B \to C + D$ with $m_C = m_A$ and $m_D = m_B$). Suppose the scattering is isotropic in the centre-of-mass frame, with differential cross section $(\mathrm{d}\sigma/\mathrm{d}\Omega)^{\rm cm} = \ell^2$ for some length scale ℓ. What is the lab-frame differential cross section for the antimuon as a function of its final energy: $(\mathrm{d}\sigma/\mathrm{d}E'_\mu)^{\rm lab}$? Compare your result for antimuon-electron scattering (for which $m_C = m_D = m_e \ll m_\mu$) and for antimuon-nucleus scattering (for which $m_C = m_D = m_N \gg m_\mu$). What is the total cross section for having the antimuon lose energy (*i.e.* for which $E'_\mu < E_\mu$ in the lab frame)? If ℓ^2 were the same for scattering with electrons and with nuclei, which type of scattering is more efficient at draining energy from an initial antimuon?

3.1.2 Particle trajectories and cross sections

We next integrate the equations of motion to determine the precise trajectory $\mathbf{r}(t)$. To this end we notice two quick integrals of the equations of motion, associated with the two conservation laws. First, because the force is conservative, taking the dot product of (3.3) with $\dot{\mathbf{r}}$ implies conservation of energy, $\dot{E} = 0$, where

$$E = \frac{m}{2}\,\dot{\mathbf{r}}^2 + U(r)\,. \tag{3.4}$$

Similarly, taking the cross product of (3.3) with $\mathbf{r}$ implies conservation of angular momentum, $\dot{\mathbf{L}} = 0$, where $\mathbf{L} = m\,\mathbf{r} \times \dot{\mathbf{r}}$.

For scattering we choose coordinates with the origin at $\mathbf{r} = 0$ and choose axes so that the relative motion is initially parallel to the y direction with speed v_i, so $\mathbf{v}_i = \dot{\mathbf{r}}(t_i) = v_i\,\mathbf{e}_y$. We choose this initial trajectory to correspond to initial motion along a line displaced relative to the y axis in the x direction by an amount b, called the impact parameter. Writing the initial position as $\mathbf{r}(t_i) = r_i\mathbf{e}_r = b\,\mathbf{e}_x + y_i\mathbf{e}_y$, the angular momentum is $\mathbf{L} = m\,\mathbf{r}_i \times \mathbf{v}_i = mbv_i\,\mathbf{e}_z$ and so points purely in the z direction, with magnitude $L = mbv_i$ that is independent of y_i. If we

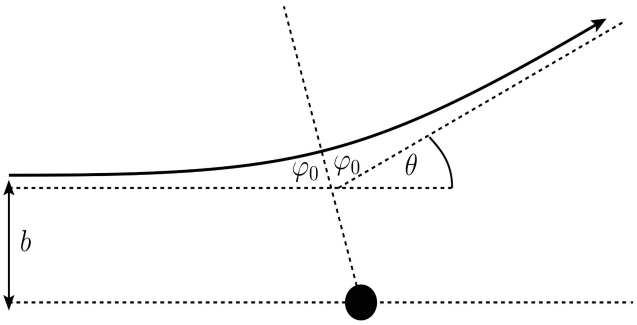


Figure 15. The geometry for scattering of a particle by a central potential. (Figure source: Itay Yavin).

choose the initial position to be infinitely far away, $y_i \to \infty$, then the total energy becomes $E = \frac{1}{2}\, m\, v_i^2$ provided we assume $U(r) \to 0$ as $r \to \infty$.

Because $\mathbf{L}$ points parallel to the z axis (and $\mathbf{L}$ is conserved) the subsequent motion can be taken to lie completely within the $x - y$ plane. The goal in a scattering problem is to compute final velocity asymptotically into the future, when $\mathbf{v} \to \mathbf{v}_f = v_f(\mathbf{e}_y \cos\theta + \mathbf{e}_x \sin\theta)$ and θ is the scattering angle in the c.o.m. frame relative to the initial direction.

To do so we first solve for the trajectories that solve the equations of motion. Using polar coordinates in the $x - y$ plane to describe $\mathbf{r}(t) = x(t)\,\mathbf{e}_x + y(t)\,\mathbf{e}_y$, with $x = r\cos\varphi$ and $y = r\sin\varphi$, the velocity is $\mathbf{v} = \dot{\mathbf{r}} = v_x\,\mathbf{e}_x + v_y\,\mathbf{e}_y$ with

$$v_x = \dot{x} = \dot{r}\,\cos\varphi - r\dot{\varphi}\,\sin\varphi \quad \text{and} \quad v_y = \dot{y} = \dot{r}\,\sin\varphi + r\dot{\varphi}\,\cos\varphi\,, \tag{3.5}$$

and so the magnitude of angular momentum is

$$L = m(x\dot{y} - y\dot{x}) = mr^2\dot{\varphi}\,, \tag{3.6}$$

while the instantaneous kinetic energy is

$$\frac{m}{2}\,\mathbf{v}^2 = \frac{m}{2}\left(\dot{x}^2 + \dot{y}^2\right) = \frac{m}{2}\left(\dot{r}^2 + r^2\dot{\varphi}^2\right) = \frac{m}{2}\,\dot{r}^2 + \frac{L^2}{2mr^2}\,. \tag{3.7}$$

The trajectories, $r(t)$ and $\varphi(t)$, are found in principle as follows. First we regard the energy equation as a first-order differential equation to be solved for $r(t)$, and the result is then used in the angular-momentum equation which is integrated to solve for $\varphi(t)$. That is, our two differential equations are

$$\frac{m\dot{r}^2}{2} + \frac{L^2}{2mr^2} + U(r) = E = \frac{mv_i^2}{2} \quad \text{and} \quad mr^2\dot{\varphi} = L = mbv_i\,, \tag{3.8}$$

and so

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \pm\sqrt{v_i^2\left(1 - \frac{b^2}{r^2}\right) - \frac{2U(r)}{m}} \quad \text{and} \quad \frac{\mathrm{d}\varphi}{\mathrm{d}t} = \frac{bv_i}{r^2}\,. \tag{3.9}$$

For scattering we are more interested in the trajectory, $r(\varphi)$, than precisely when we arrive at any one point on this trajectory as a function of time, and so can take the ratio of the above two equations to get

$$\frac{\mathrm{d}r}{\mathrm{d}\varphi} = \frac{\dot{r}}{\dot{\varphi}} = \pm\sqrt{\frac{r^2}{b^2}(r^2 - b^2) - \frac{2r^4 U(r)}{mb^2 v_i^2}}\,. \tag{3.10}$$

These formulae show what Fig. 15 also indicates: the radial distance, r, initially decreases (so $\dot{r} < 0$) until the trajectory reaches its point of closest approach, where $\dot{r} = 0$ instantaneously. After this point $\dot{r} > 0$ indicating that the radial distance grows with time. As also shown in the figure, the point of closest approach occurs at the angular position φ_0 and this is precisely half of the total change in φ as t sweeps from $-\infty$ to ∞. The scattering angle, θ, is therefore related to φ_0 by the relation

$$2\varphi_0 + \theta = \pi\,. \tag{3.11}$$

The strategy is first to compute φ_0 and then use (3.11) to obtain θ.

Since the point of closest approach satisfies $\dot{r} = 0$ it must occur at a radius, r_0, that satisfies

$$\frac{b^2}{r_0^2} = 1 - \frac{2U(r_0)}{mv_i^2}\,, \tag{3.12}$$

and so $r_0 = b$ in the absence of interactions (as would be expected for a straight-line trajectory). The radius r_0 determines φ_0 through the formula of the orbit, $r_0 = r(\varphi_0)$, where $r(\varphi)$ satisfies (3.10), so

$$\varphi_0 = \int_{r_0}^{\infty} \frac{\mathrm{d}r}{\mathrm{d}r/\mathrm{d}\varphi} = b\int_{r_0}^{\infty} \frac{\mathrm{d}r}{r\sqrt{r^2 - b^2 - \frac{2r^2U(r)}{mv_i^2}}}\,. \tag{3.13}$$

This can be obtained in principle by integrating once $U(r)$ is known, after which the scattering angle is $\theta = \pi - 2\varphi_0$, and for fixed potential and v_i dictates $\theta = \theta(b)$. In principle this can be inverted to learn $b = b(\theta)$ for these trajectories.

The utility of having $b(\theta)$ is that this is what governs the differential scattering rate if an ensemble of particles all sharing the same speed approaches the target with a range of impact parameters, b. That is, suppose we are given a uniform luminosity, $\mathcal{L} = n_B v_i$, of incident particles evenly distributed in impact parameter. Then a number $\mathrm{d}N = (2\pi b\,\mathrm{d}b)\mathcal{L}$ of these lie in an infinitesimal range $\mathrm{d}b$ about any particular impact parameter b. All of these particles emerge at late time within a range $\mathrm{d}\theta$ around the scattering angle $\theta(b)$ dictated by the particle trajectories found above, and so

$$\mathrm{d}N = 2\pi b\,\mathcal{L}\left|\frac{\mathrm{d}b}{\mathrm{d}\theta}\right|\mathrm{d}\theta\,, \tag{3.14}$$

and so the differential cross section is $\mathrm{d}\sigma = \mathrm{d}N/\mathcal{L}$ and is given by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\theta} = 2\pi b(\theta)\left|\frac{\mathrm{d}b}{\mathrm{d}\theta}\right| \qquad \text{or} \qquad \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{b(\theta)}{\sin\theta}\left|\frac{\mathrm{d}b}{\mathrm{d}\theta}\right| . \tag{3.15}$$

Clearly knowledge of the potential $U(r)$ allows $b(\theta)$ — and hence $\mathrm{d}\sigma/\mathrm{d}\theta$ — to be calculated.

3.1.3 Scattering from a hard sphere

The simplest case is the case of a hard sphere, for which $U(r) = 0$ for $r > R$ and $U = \infty$ for $r < R$. In this case an incoming particle experiences a purely normal force at the sphere's surface that requires the sign of the radial component of velocity to instantaneously change sign without affecting the tangential component. Energy conservation then requires the reflected radial component to have precisely the same magnitude as it did before reflection. Together these imply the trajectory reflects off the sphere's surface, departing with an angle to the surface normal that is the same size as the angle it had to the normal when it came in (see Figure 16).

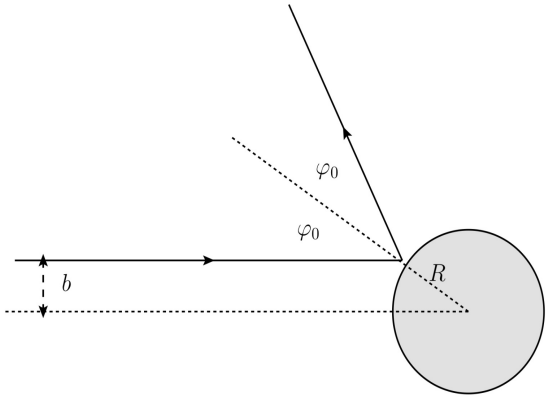


Figure 16. The geometry for scattering from a hard sphere. (Figure source: Itay Yavin).

We first seek the angular position, φ_0, of closest approach and in this case it is simply the position where the trajectory meets the sphere. When $b > R$ the trajectory misses the sphere and so remains a straight line, and consequently $\varphi_0 = \frac{\pi}{2}$ (which means $\theta = \pi - 2\varphi_0 = 0$, so no scattering). But if $b \leq R$ then the geometry of Figure 16 shows that φ_0 is related to b and R by

$$b = R\,\sin\varphi_0 = R\cos\frac{\theta}{2}\,, \tag{3.16}$$

and the second equality again trades φ_0 for θ using (3.11). From this we see $|\mathrm{d}b/\mathrm{d}\theta| = \frac{1}{2}\,R\,\sin\frac{\theta}{2}$ and so

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{R\cos(\theta/2)}{\sin\theta}\left[\frac{R\,\sin(\theta/2)}{2}\right] = \frac{R^2}{4}\,, \tag{3.17}$$

which uses the multiple-angle formula $\sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$.

We find the scattering is *isotropic*: because the right-hand-side is independent of θ there is equal differential likelihood to scatter into any particular angular direction. The total cross section is found by integrating $\mathrm{d}\sigma/\mathrm{d}\Omega$ over the 4π solid angle, and so gives

$$\sigma_{\text{tot}} = \int \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\,\mathrm{d}\Omega = \pi R^2\,, \tag{3.18}$$

and so agrees with the sphere's geometric cross section (as one might have guessed).

3.1.4 Rutherford scattering

The classical scattering of two point charges due to their Coulomb interaction is called *Rutherford* scattering, and is the result to which Rutherford compared his scattering measurements when discovering the nucleus.

To obtain the cross section in this case we specialize the above discussion to $U(r) = \kappa_c/r$ where $\kappa_c = q_A q_B$ is the product of charges in Gaussian units (or $\kappa_c = q_A q_B/4\pi\epsilon_0$ in SI units). In this case the condition (3.12) that fixes the radius, r_0, of closest approach becomes

$$\frac{b^2}{r_0^2} = 1 - \frac{2\kappa_c}{m r_0 v_i^2} \quad \text{which inverts to} \quad \frac{1}{r_0} = \frac{1}{b}\Big(\sqrt{1+x^2} - x\Big) \quad \text{for} \quad x := \frac{\kappa_c}{m v_i^2 b}\,. \tag{3.19}$$

The dimensionless quantity x is the ratio of Coulomb energy at distance b to the initial kinetic energy, and so is a measure of the importance of the Coulomb interaction for the scattering event (with $r_0 \to b$, as appropriate for a straight trajectory, as $x \to 0$).

To obtain the angular position, φ_0, of closest approach we perform the integration in (3.13), which for this potential can be done in closed form

$$\varphi_0 = b\int_{r_0}^{\infty} \frac{\mathrm{d}r}{r\sqrt{r^2 - b^2 - \frac{2\kappa_c r}{m v_i^2}}} = \int_0^{b/r_0} \frac{\mathrm{d}u}{\sqrt{1 - 2xu - u^2}} = \arccos\left(\frac{x}{\sqrt{1+x^2}}\right)\,, \tag{3.20}$$

which uses the change of integration variables $u = b/r$. Notice $\varphi_0 \to \frac{\pi}{2}$ when $x \to 0$, as it should, although $\varphi_0 \neq \frac{\pi}{2}$ for any finite b, no matter how large. This is a reflection of the extremely long range of the Coulomb interaction.

Inverting gives $x = \cot\varphi_0$ and so

$$b = \frac{\kappa_c}{m v_i^2}\,\tan\varphi_0 = \frac{\kappa_c}{m v_i^2}\,\cot\frac{\theta}{2}\,, \tag{3.21}$$

using (3.11) to trade φ_0 for θ. Using this expression for $b(\theta)$ in the differential cross section then gives the standard result

$$\boxed{\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left(\frac{\kappa_c}{2m v_i^2}\right)^2 \csc^4\frac{\theta}{2}}\,. \tag{3.22}$$

Exercise 3.3: Repeat the arguments used for Rutherford scattering to calculate the classical centre-of-mass scattering angle as a function of impact parameter, $\theta(b)$, for two particles with initial relative speed v_i interacting through the potential $U(r) = \kappa/r^2$, with $\kappa > 0$. Invert this to obtain $b(\theta)$ and use the answer to compute the differential cross section, $\mathrm{d}\sigma/\mathrm{d}\Omega$. Is the small-angle scattering you find stronger or weaker than for Rutherford scattering? Why might this be so? Suppose the attraction were attractive ($\kappa < 0$) with $|\kappa| > \frac{1}{2}mv_i^2$. Do you see any problem computing $\theta(b)$ in this case?

Several things are noteworthy about the Rutherford expression (3.22).

- First off, because it depends on κ_c^2 the result for $\mathrm{d}\sigma/\mathrm{d}\Omega$ does not depend on the relative sign of q_A and q_B. This is because the scattering trajectories are hyperbolae for either sign, and it does not matter for the cross section whether the particle trajectories are deflected towards or away from one another so long as the deflection angle for a given b is the same.

- Second, the incidence of very large scattering angles can be high, as Rutherford noticed. Integrating (3.22) through a range of angles $\theta_{\min} \leq \theta \leq \pi$, we find

$$\sigma(\theta > \theta_{\min}) = 2\pi \int_{\theta_{\min}}^{\pi} \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right) \sin\theta \mathrm{d}\theta = \pi \left(\frac{\kappa_c}{mv_i^2}\right)^2 \cot^2 \frac{\theta_{\min}}{2}\,, \tag{3.23}$$

  which becomes $\pi(\kappa_c/mv_i^2)^2$ when $\theta_{\min} = \frac{\pi}{2}$, as appropriate for the hemisphere where the incident particle back-scatters in the c.o.m. frame. This is like the area of a sphere whose radius is κ_c/mv_i^2, and this length scale can be much larger than the physical size of the individual charges because the Coulomb force has such a long range. A natural choice for the lower limit is set by the width of the incident beam of particles, since this imposes a maximum impact parameter. For Rutherford scattering (3.21) implies $\cot(\theta_{\min}/2) = mv^2 b_{\max}/\kappa_c$, and so

$$\sigma(b < b_{\max}) = \pi b_{\max}^2\,. \tag{3.24}$$

- Third, although not realistic in practice (since beams are not infinitely large), notice that in principle the integrated cross section diverges if it is integrated right down to $\theta_{\min} \to 0$. This arises because small-angle scattering corresponds to large impact parameter, b, and its divergence reflects the fact that there is small but nonzero scattering even for arbitrarily large b. Again this is the Coulomb interaction's long range at work, though in practical settings one of our approximations will really fail before we get out to arbitrarily large distances. Sometimes this is because of the finite beam size, but it also can happen that the charge is screened at large distances by competition

with other particles with opposite charges that are attracted to the Coulomb source. This is what happens for low-energy α particles scattering from atomic nuclei, since for impact parameters larger than the atomic size the atomic electrons can screen the nuclear charge.

3.1.5 Light deflection by the Sun

Calculations of scattering with a $1/r$ potential also apply when the interaction at play is Newton's law of gravity rather than the Coulomb law, although for gravity systems are so big that one does not normally have beams of particles comparable to the size of the scatterer.

The above expression for $\theta(b)$ does apply to the motion of individual particles in a central gravitational field, however, once we make the replacement $\kappa_c \to \kappa_g = Gm_Am_B$. Keeping in mind that $m = m_Am_B/(m_A+m_B)$ is the reduced mass, we see that for gravitational scattering

$$\tan\frac{\theta}{2} = \frac{Gm_Am_B}{mv_i^2 b} = \frac{G(m_A+m_B)}{v_i^2 b} \ . \tag{3.25}$$

In the special case of light moving past the Sun we can take $v = c$ as well as $m_A + m_B \simeq M_\odot$ to be the solar mass, in which case we find $\tan\frac{\theta}{2} \simeq \frac{\theta}{2} \simeq GM_\odot/bc^2$, where the small-angle expansion of $\tan\frac{\theta}{2}$ is justified by the small size of $GM_\odot/bc^2$. For instance, taking $b \sim R_\odot \simeq 700,000$ km and using $2GM_\odot/c^2 \simeq 3.0$ km gives $\theta \simeq 2GM_\odot/bc^2 \simeq 4.3 \times 10^{-6}$ radians, or 0.9 seconds of arc.

Of course it is suspicious to apply Newtonian formulae to relativistic systems, for which Einstein provided the first proper calculation of light deflection and obtained a result twice as large: $\theta \simeq 4GM_\odot/bc^2$. Observations of the deflection of starlight near the Sun (which become visible during a solar eclipse, for instance) agree with Einstein's value, and this agreement was partly what led to the acceptance of his theory of gravity: the General Theory of Relativity.

3.1.6 Impulse approximation

In principle the above calculations provide a definitive answer to the question of how particles scatter classically when they interact through a central conservative force. Although we got lucky with the Rutherford problem which could be solved in closed form, in general the determination of r_0 and φ_0 must only be approximate. One way to do so is to evaluate them numerically, in which case the approximation can be very good. But it is also useful to have analytic approximations, both to check numerics but also to be able to explore dependence on parameters when new kinds of interactions are considered.

One such an approximation is the *impulse* approximation, which applies when the interactions are weak and so the scattering angles are small. Besides being useful in its own right, discussing it here also sets up a similar technique that is useful when we consider quantum scattering.

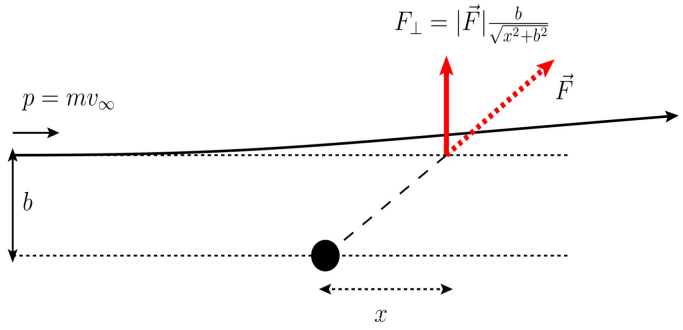


Figure 17. The geometry for the impulse approximation. (Figure source: Itay Yavin).

When the interaction is weak the real trajectory taken by a particle is not much different from a straight line. The impulse approximation starts from the observation that the net momentum transfer to a particle is given by $\Delta\mathbf{p} = \int \mathrm{d}t\, \dot{\mathbf{p}} = \int \mathrm{d}t\, \mathbf{F}$, and it is the component of $\Delta\mathbf{p}$ that is transverse to the initial momentum, $\mathbf{p}_i$, that governs the deflection angle (while forces parallel to the trajectory speed the particle up and slow it down, rather than deflecting its trajectory).

Of course doing this integral is hard, partly because it requires knowing the detailed trajectory in order to compute the applied force. The impulse approximation side-steps this complication by taking the trajectory at leading order to simply be the straight line that would have been taken without the application of the force (see Figure 17):

$$\Delta p_\perp = \int_{-\infty}^{\infty} \mathrm{d}t\, F_\perp = \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{v_i}\, F_\perp = \int_{-\infty}^{\infty} \frac{\mathrm{d}x}{v_i}\, F \sin\varphi = \frac{b}{v_i}\int_{-\infty}^{\infty} \mathrm{d}x\, \frac{F}{\sqrt{x^2+b^2}}\,, \tag{3.26}$$

with the scattering angle then computed using

$$\theta \simeq \tan\theta \simeq \frac{\Delta p_\perp}{p_i} = \frac{b}{mv_i^2}\int_{-\infty}^{\infty} \mathrm{d}x\, \frac{F}{\sqrt{x^2+b^2}}\,. \tag{3.27}$$

For example, as applied to the Coulomb force we have $F = \kappa_c/(x^2+b^2)$ and so

$$\theta \simeq \frac{\kappa_c}{mv_i^2 b}\int_{-\infty}^{\infty} \frac{\mathrm{d}u}{(u^2+1)^{3/2}} = \frac{2\kappa_c}{mv_i^2 b}\,. \tag{3.28}$$

This indeed agrees with the small-angle expansion of the full Coulomb result, (3.21), which says $\tan(\theta/2) = \kappa_c/(mv_i^2 b)$. This result also better quantifies precisely when the impulse approximation works, because it says that the scattering is through small angles (the regime for which the impulse approximation is justified) only if $\kappa_c/b \ll mv_i^2$; *i.e.* if the impact parameter is large enough that the Coulomb energy at closest approach is much smaller than the initial kinetic energy.

Less trivially, suppose one of the two particles discussed to this point actually consists of two particles that are a distance $a \ll b$ from one another and carry equal and opposite charges. Suppose in particular these two sub-charges are displaced from one another by a distance a along the y axis, with charge q_B situated at $y = +a/2$ and charge $-q_B$ located at $y = -a/2$. Then, for $a \ll b$, the impulse approximation predicts

$$F_{\perp,\pm} = \pm \frac{\kappa_c}{x^2 + (b \mp a/2)^2} \frac{b}{\sqrt{x^2+b^2}} \simeq \pm \frac{\kappa_c b}{(x^2+b^2)^{3/2}} \left[1 \pm \frac{ab}{x^2+b^2} - \frac{a^2(x^2-3b^2)}{4(x^2+b^2)^2} + \cdots \right], \tag{3.29}$$

and so these sum to give

$$F_\perp = \frac{\kappa_c a b^2}{(x^2+b^2)^{5/2}} + \cdots . \tag{3.30}$$

The leading contribution to the scattering angle then becomes

$$\theta \simeq \frac{\kappa_c a}{m v_i^2 b^2} \int_{-\infty}^{\infty} \frac{\mathrm{d}u}{(u^2+1)^{5/2}} = \frac{4\kappa_c a}{3 m v_i^2 b^2} . \tag{3.31}$$

This predicts a small-angle cross section that as $\theta \to 0$ is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq \frac{2\kappa_c a}{3 m v_i^2 \theta^3} . \tag{3.32}$$

This varies less strongly than the $1/\theta^4$ of the Rutherford result, and does so because the net force falls off more quickly than does the Coulomb interaction due to the two source particles having opposite charges (so their far-field Coulomb forces cancel). What survives dominantly at large distances is the dipole interaction, for which the potential falls off like $1/r^2$ rather than $1/r$, and because small-angle scattering occurs at large b this faster falloff translates into reduced small-angle scattering.

Exercise 3.4: For a conservative, central force, where $\mathbf{F} = -\nabla U$ for $U = U(r)$ and $r = |\mathbf{r}|$, show that the impulse approximation predicts

$$\theta \simeq -\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{r} \frac{\mathrm{d}U}{\mathrm{d}r} = -\int_{-\infty}^{\infty} \frac{\mathrm{d}x}{x} \frac{\mathrm{d}U}{\mathrm{d}x} , \tag{3.33}$$

where $r = \sqrt{x^2+b^2}$. It is tempting to integrate this by parts and use $U \to 0$ as $r \to \infty$ to derive $\theta \simeq -\int_{-\infty}^{\infty} \mathrm{d}x (U/x^2)$. Does this last result give the correct answer when $U = \kappa/r$? (Does its sign agree with (3.33) in this case?) If not, what went wrong in the integration-by-parts argument?

Exercise 3.5: Use the impulse approximation to compute the scattering angle $\theta(b)$ and differential cross section $\mathrm{d}\sigma/\mathrm{d}\Omega$ for scattering from a conservative, central potential of the form $U(r) = \kappa_U / r^n$ where $r = |\mathbf{r}|$ and $n \geq 1$.

3.2 Quantum potential scattering

We next turn to the calculation of scattering processes using non-relativistic quantum mechanics. The arguments parallel the developments given above for classical scattering. In this section the connection is made between the cross section and the general properties of the quantum wave-function. The next section then addresses how to compute this wave-function given the properties of an interaction potential.

The starting point for quantum systems is the time-dependent Schrödinger equation, since this governs how systems evolve in time. Our interest is in particular in the evolution of the two-particle wave-function, $\Psi(\mathbf{r}_A, \mathbf{r}_B, t)$, describing two scattering particles that interact with one another through a central potential, $U(r)$, with $r = |\mathbf{r}_A - \mathbf{r}_B|$. To start with we ignore any possible internal quantum numbers (such as electronic energy levels for an atom) from which energy can be extracted (or deposited) and so concentrate for the time being on elastic scattering.

The system of interest is therefore

$$i\frac{\partial \Psi}{\partial t} = -\frac{1}{2m_A}\nabla_A^2\Psi - \frac{1}{2m_B}\nabla_B^2\Psi + U\,\Psi\,, \tag{3.34}$$

from which we seek to predict the state at late times, $\Psi(t \to \infty)$, given initial conditions $\Psi(t \to -\infty)$. Here ∇_A is the usual gradient with respect to $\mathbf{r}_A$, and ∇_B is its counterpart for $\mathbf{r}_B$.

3.2.1 The equivalent one-body problem

We start by converting the two-body scattering problem into a one-body problem, by isolating the centre of mass. To this end define as before $\mathbf{R} = (m_A\mathbf{r}_A + m_B\mathbf{r}_B)/(m_A + m_B)$ and $\mathbf{r} = \mathbf{r}_A - \mathbf{r}_B$, and change variables from $\Psi(\mathbf{r}_A, \mathbf{r}_B, t)$ to $\Psi(\mathbf{R}, \mathbf{r}, t)$ in the Schrödinger equation, to get

$$i\frac{\partial \Psi}{\partial t} = -\frac{1}{2M}\nabla_R^2\Psi - \frac{1}{2m}\nabla^2\Psi + U(r)\,\Psi\,, \tag{3.35}$$

where $M := m_A + m_B$ is the total mass and $m = m_A m_B/M$ is the reduced mass. Here ∇_R denotes the gradient with respect to $\mathbf{R}$ and ∇ represents the same for $\mathbf{r}$.

This now has terms involving only $\mathbf{R}$ and those involving only $\mathbf{r}$, and this reflects in the quantum theory how these evolve independently, just as was true for the classical theory. We can therefore choose our state to be a product state: $\Psi(\mathbf{R}, \mathbf{r}, t) = \chi(\mathbf{R}, t)\,\psi(\mathbf{r}, t)$, so that their probabilities are initially uncorrelated. Once this is true for an initial time the same remains true for later times since the Schrödinger equation becomes equivalent to the pair of equations

$$\begin{aligned} i\frac{\partial \chi}{\partial t} + \frac{1}{2M}\nabla_R^2\chi &= \lambda\,\chi \\ i\frac{\partial \psi}{\partial t} + \frac{1}{2m}\nabla^2\psi - U(r)\,\psi &= -\lambda\psi\,, \end{aligned} \tag{3.36}$$

for λ an arbitrary constant. The first of these describes the free-particle motion of the overall c.o.m. while the second describes single-particle Schrödinger equation for motion in the presence of the potential U. The constant λ amounts to the freedom to choose our zero of energy for either the $\mathbf{R}$ system or the $\mathbf{r}$ system but not both, and we use this freedom to set $\lambda = 0$ for the ψ equation. We arrive in this way to the equation governing the relative motion of the two particles, whose solutions we wish to study in more detail:

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m}\nabla^2\psi + U(r)\,\psi\,. \tag{3.37}$$

Also of interest is its time-independent analog, satisfied by energy eigenstates whose time-dependence is particularly simple: $\psi(\mathbf{r},t) = \psi(\mathbf{r})\,e^{-iEt}$:

$$-\frac{1}{2m}\nabla^2\psi + U(r)\,\psi = E\psi\,. \tag{3.38}$$

Finally, we assume $U(r) \to 0$ for large r so that the interactions turns off when particles are well-separated. This means that for very large r we have approximate solutions to (3.38) of the plane-wave form

$$\psi_E(\mathbf{r},t) \propto e^{-i(Et-\mathbf{k}\cdot\mathbf{r})} \qquad (\text{for large } r)\,, \tag{3.39}$$

where $k^2 := \mathbf{k}\cdot\mathbf{k} = 2mE$ but the direction of $\mathbf{k}$ is arbitrary. Of course any linear combination of these solutions is also an allowed energy eigenstate (for sufficiently large r).

3.2.2 Time-dependent *vs* time-independent

Since scattering involves nontrivial time evolution one might think that really only the time-dependent Schrödinger equation should be of interest. After all, energy eigenstates do not evolve nontrivially in time at all (and momentum eigenstates have the same probability to be everywhere), so how can they describe something as temporal as scattering? This section argues that this reasoning need not preclude using the time-independent equation for scattering problems. (Those needing no convincing on this point should skip directly to the scattering boundary condition, given in eq. (3.44).)

It can do so because of the specific nature of a scattering problem. For this we start with particles that do not much interact initially, usually because they are too far apart. Scattering happens because we bring these particles much closer together, but only briefly. They do something interesting because some interaction is temporarily important, but then turns off again as the final-state particles again separate. It is true that one way to approach this is to take our initial states as wave packets (and so chosen not to be exact momentum or energy eigenstates, though not so narrowly that the uncertainty relations preclude our assigning the packet both position and momentum to within the experimental accuracy). We would then let these packets evolve using the time-dependent equation, (3.37), and solve for the subsequent evolution into separating wave-packets. But this is not the only way.

Instead we adopt what is an equivalent description, but one that allows us to use energy eigenstates and so instead use (3.38). This approach is based on the observation that eigenstates with $E > U(\infty)$ are usually degenerate (since E doesn't care about the direction of the momentum) and so boundary conditions are not usually completely specified just by normalization conditions (in the same way they are for bound states, say). Consequently these boundary conditions can be used to formulate the scattering problem using energy eigenstates.

To see what this means, recall how things work for scattering of a particle by a square well in first exposures to single-variable quantum mechanics. In regions where the potential is constant, $U = U_0$, the energy eigenstates are degenerate because both e^{ikx} and e^{-ikx} have the same energy: $E = U_0 + k^2/2m$. For bound states within the square well the same is not true because these have energy $E < U_0$ at large x, and so the eigenstates are instead e^{kx} and e^{-kx}. The degeneracy is then broken (and the energy fixed to a quantized value) by the requirement that the states do not grow exponentially as $x \to \pm\infty$. But the same conditions are not available for states with $E > U$ at infinity, since in this case neither of e^{ikx} and e^{-ikx} have better normalization properties than the other. Instead, for particles approaching the potential from the left (say) the boundary conditions are normally chosen to include both incoming and outgoing (reflected) waves — *i.e.* $\psi \propto e^{-i(Et-kx)} + R\, e^{-i(Et+kx)}$ — to the left of the potential, but with only outgoing waves (with no ingoing wave) — *i.e.* $\psi \propto T e^{-i(Et-kx)}$ — to the right. Determining the unknown coefficients R and T is the core of a scattering problem formulated this way, since these respectively give the amplitudes for reflection from and transmission through the potential. Energy eigenstates chosen in this way are called *scattering eigenstates*.

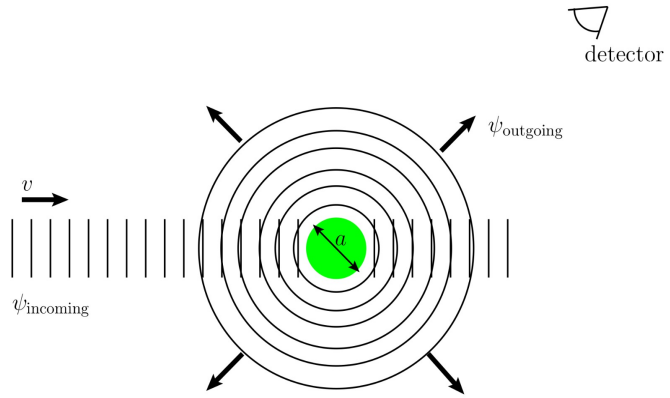


Figure 18. The geometry for scattering in the Schrödinger equation. (Figure source: Itay Yavin).

For scattering in three dimensions we do a similar thing. In this case — see Figure (18)

— we imagine the incoming particles to arrive from the negative z axis (say) before impinging on the scattering potential. This corresponds to a solution to (3.38) with $U = 0$ of the form:

$$\psi_i = C\, e^{-i(Et-kz)}\,, \tag{3.40}$$

where $k^2 = 2mE$ and C is a normalization constant whose value doesn't matter in what follows.

We expect the outgoing wave, after scattering, in this case should head out radially in all directions (possibly with an amplitude modulated with direction) as shown in Figure 18. More precisely, for $r \to \infty$ the asymptotic solution should have a component looking like an outgoing spherical wave. In the absence of a potential the Schrödinger equation in spherical coordinates is

$$-\nabla^2\psi = -\left[\frac{\partial^2\psi}{\partial r^2} + \frac{2}{r}\,\frac{\partial\psi}{\partial r} + \frac{1}{r^2}\,\Delta\psi\right] = 2mE\,\psi\,, \tag{3.41}$$

where Δ is the following differential operator that depends only on the angular variables:

$$\Delta\psi = \frac{1}{\sin\theta}\,\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\psi}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\,\frac{\partial^2\psi}{\partial\phi^2}\,. \tag{3.42}$$

Spherical waves are approximate solutions to this equation of the form

$$\psi_{\rm out}(r,t) \propto \frac{e^{-i(Et-kr)}}{r} \quad \text{(out-going)} \quad \text{or} \quad \psi_{\rm in}(r,t) \propto \frac{e^{-i(Et+kr)}}{r} \quad \text{(in-going)} \tag{3.43}$$

where (again) $k^2 = 2mE$. These solutions are approximate inasmuch as they fail[18] to solve (3.41) only by terms that are subdominant in powers of $1/r$. We can also allow an angle-dependent normalization factor, $w(\theta,\phi)$, and still solve the equation for large enough r because the angular piece of (3.41) is subdominant at large r (more about this below).

Combining both the incoming and outgoing waves, the scattering boundary condition we seek for large r is a linear combination of the incoming plane wave and an outgoing spherical wave. Consequently we ask ψ at large distances to have the form

$$\psi(r,\theta,\phi,t) \to C\left[e^{-i(Et-kz)} + w(\theta,\phi)\,\frac{e^{i(kr-Et)}}{r}\right] \qquad \text{(for very large } r\text{)}\,. \tag{3.44}$$

The explicit form for $w(\theta,\phi)$ depends on the precise form of $U(r)$ that is responsible for the scattering. It is $w(\theta,\phi)$ that directly carries the information about the interaction out to spatial infinity and so determines the cross section, as is now shown explicitly.

[18] As we see below the exact solutions to (3.41) involve spherical Bessel functions, which look like linear combinations of incoming and outgoing spherical waves asymptotically as $r \to \infty$.

3.2.3 Cross section and scattering amplitude

To make the connection between w and $\mathrm{d}\sigma$ recall that the probability density (probability per unit volume) carried by a wave-function is $\rho(\mathbf{r},t) = \psi^*\psi$ while the probability *flux* (probability flow per unit area per unit time) is

$$\mathbf{j}(\mathbf{r},t) = \frac{i}{2m}\left(\psi\nabla\psi^* - \psi^*\nabla\psi\right) . \tag{3.45}$$

This latter is called the probability flux because the Schrödinger equation implies the probability density is conserved inasmuch as they are related by $\partial\rho/\partial t + \nabla\cdot\mathbf{j} = 0$. This expresses conservation of probability because the only way the probability, $P(\mathcal{R}) = \int_{\mathcal{R}} \mathrm{d}^3x\ \rho$, of a particle being within a region, $\mathcal{R}$, can change is by physically moving probability out through the surface, $\partial\mathcal{R}$, that marks it boundary, via the probability flux, $\mathbf{j}$. That is:

$$\frac{\mathrm{d}P(\mathcal{R})}{\mathrm{d}t} = \int_{\mathcal{R}} \mathrm{d}^3x\ \frac{\partial\rho}{\partial t} = -\int_{\mathcal{R}} \mathrm{d}^3x\ \nabla\cdot\mathbf{j} = -\oint_{\partial\mathcal{R}} \mathrm{d}^2x\ \mathbf{n}\cdot\mathbf{j}\,, \tag{3.46}$$

where d^2x denotes a differential surface area on the boundary, $\partial\mathcal{R}$, and $\mathbf{n}$ is its outward-pointing normal. The last equality uses Stokes' theorem to relate the volume integral over $\nabla\cdot\mathbf{j}$ to the surface flux: *i.e.* the surface integral of $\mathbf{n}\cdot\mathbf{j}$.

> **Exercise 3.6:** Prove that the time-dependent Schrödinger equation for ψ implies that $\rho = \psi^*\psi$ and $\mathbf{j} = (i/2m)(\psi\nabla\psi^* - \psi^*\nabla\psi)$ are related by $\partial\rho/\partial t + \nabla\cdot\mathbf{j} = 0$. This equation can also be written as $\partial_\mu j^\mu = 0$ where $j^\mu = \{\rho, \mathbf{j}\}$ is a 4-vector. Use the Lorentz-transformation rule for a 4-vector to calculate how ρ and $\mathbf{j}$ are related for observers in two frames that move relative to one another with speed v. Prove also that $\partial_\mu j^\mu$ is a Lorentz invariant quantity.

Applied to the incoming wave, $\psi_i = C\, e^{-i(Et-kz)}$, the probability density and flux become $\rho = |C|^2$ and

$$\mathbf{j} = -\frac{i}{2m}\left(2ik\mathbf{e}_z\right)\psi_i^*\psi_i = \frac{\rho k}{m}\,\mathbf{e}_z = \rho\, v_i\,\mathbf{e}_z = |C|^2 v_i\,\mathbf{e}_z\,. \tag{3.47}$$

From this we see that for N such particles the average density of particles is $n_B = N\rho$ and the average particle flux of initial particles — or luminosity — is

$$\mathcal{L} = N\mathbf{j}\cdot\mathbf{e}_z = N\rho v_i = n_B v_i = N|C|^2 v_i\,. \tag{3.48}$$

Similarly, applied to the out-going spherical wave at large r we have

$$\begin{aligned}\mathbf{j} &= -\frac{i}{2m}\left(2ik\,\mathbf{e}_r\right)\frac{|C|^2|w(\theta,\phi)|^2}{r^2} + (\text{terms falling faster than } 1/r^2)\\ &= \frac{|C|^2|w(\theta,\phi)|^2 v_i}{r^2}\,\mathbf{e}_r + (\text{terms falling faster than } 1/r^2)\,,\end{aligned} \tag{3.49}$$

and so points radially, up to terms falling faster than $1/r^2$. The final equality uses the elasticity of the scattering to conclude that k is the same as for the initial state and so we can again use $k/m = v_i$.

For N particles the rate with which particles pass through a surface element subtending solid angle $\mathrm{d}\Omega$ at a large distance r whose normal, $\mathbf{n} = \mathbf{e}_r$, points in the radial direction labelled by (θ, ϕ) is

$$\mathrm{d}N = \lim_{r\to\infty} N\,\mathbf{j}\cdot\mathbf{n}(r^2\mathrm{d}\Omega) = N|C|^2 v_i |w(\theta,\phi)|^2 \mathrm{d}\Omega\,, \tag{3.50}$$

where $r^2\mathrm{d}\Omega$ is the area of the surface element. Dividing this by the luminosity gives the differential cross section, $\mathrm{d}\sigma$, and so we arrive at the desired result

$$\boxed{\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{1}{\mathcal{L}}\frac{\mathrm{d}N}{\mathrm{d}\Omega} = |w(\theta,\phi)|^2}\,. \tag{3.51}$$

$w(\theta,\phi)$ is the *scattering amplitude*, whose square determines the scattering cross section.

For the purposes of scattering everything comes down to computing $w(\theta,\phi)$. In principle this is done by explicitly solving (3.38) and fixing the integration constants by requiring agreement with the asymptotic form (3.44) at large r. In practice this must often be done numerically, though it is possible to solve explicitly in closed form for some special cases like the scattering from a hard sphere or for the Coulomb potential, as we describe in the next section. More generally we require an approximation scheme, several of which are described in the following sections.

3.2.4 Partial waves

To solve for $w(\theta,\phi)$ we start by partially solving (3.38) in spherical polar coordinates, to more precisely pin down its angular dependence. For a central potential, $U = U(r)$, angular momentum conservation plays an important role in describing the angular dependence for classical scattering and the same is also true for quantum scattering from a central potential.

The general scattering solution, ψ, to the time-independent Schrödinger equation can be written as a linear combination, $\psi = \sum_{\ell\ell_z} C_{\ell\ell_z}\,\psi_{\ell\ell_z}$, of a *basis*[19] of solutions that have a separated form: $\psi_{\ell\ell_z}(k;r,\theta,\phi) = R_\ell(kr)\,Y_{\ell\ell_z}(\theta,\phi)$, where the functions R_ℓ and $Y_{\ell\ell_z}$ satisfy the radial equation

$$\frac{\mathrm{d}^2R_\ell}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}R_\ell}{\mathrm{d}r} + \left[2mE - \frac{\ell(\ell+1)}{r^2} - 2mU(r)\right]R_\ell = 0\,, \tag{3.52}$$

and angular equations

$$-\Delta Y_{\ell\ell_z} = \ell(\ell+1)\,Y_{\ell\ell_z} \qquad \text{and} \qquad \frac{\partial Y_{\ell\ell_z}}{\partial\phi} = i\ell_z Y_{\ell\ell_z}\,. \tag{3.53}$$

[19] And if this all sounds like expanding a vector in a complete set of basis vectors in linear algebra, it should!

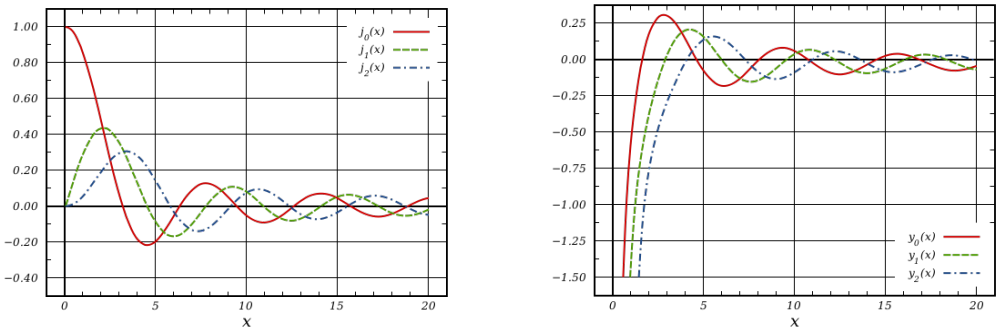
Here $\ell = 0, 1, 2, \cdots$ is the total angular-momentum quantum number, with $\mathbf{L}^2\psi_{\ell\ell_z} = -\Delta\psi_{\ell\ell_z} = \ell(\ell+1)\psi_{\ell\ell_z}$, while $\ell_z \in \{-\ell, -\ell+1, \cdots, \ell-1, \ell\}$ is the 'magnetic' quantum number that gives the eigenvalue of the z-component of $\mathbf{L}$: $L_z\psi_{\ell\ell_z} = -i\partial\psi_{\ell\ell_z}/\partial\phi = \ell_z\psi_{\ell\ell_z}$.

In practice we are interested in solutions with $\ell_z = 0$, because both $\mathbf{L}$ and L_z are conserved (for central potentials, $U(r)$) and acting with L_z on the initial state gives

$$L_z\psi_i = L_z\left[C\,e^{-i(Et-kz)}\right] = L_z\left[C\,e^{-i(Et-kr\cos\theta)}\right] = 0\,. \tag{3.54}$$

For $\ell_z = 0$ the angular wave-function, $Y_{\ell\ell_z}(\theta,\phi)$, simplifies to a Legendre polynomial (an order-ℓ polynomial in $\cos\theta$):

$$Y_{\ell\,0} \propto P_\ell(\cos\theta)\,. \tag{3.55}$$



(a) $j_\ell(x)$ (b) $y_\ell(x)$

Figure 19. Plots of the first few spherical Bessel functions. (Figure source: Wikipedia "https://en.wikipedia.org/wiki/Bessel_function").

Let's now focus on regions where $U \to 0$, which we assume includes the regime at very large r, and so consider solutions to the Schrödinger equation when U vanishes. General solutions to the radial equation, (3.52), are given for $U = 0$ by[20]

$$R_\ell = A_\ell\, j_\ell(kr) + B_\ell\, y_\ell(kr)\,, \tag{3.56}$$

where $k = \sqrt{2mE}$, A_ℓ and B_ℓ are integration constants and $j_\ell(x)$ and $y_\ell(x)$ are spherical Bessel functions (with $j_\ell(x)$ the ones that are finite as $x \to 0$), the first few of which are plotted in Figure 19. They can be written as order-ℓ polynomials in (trig functions)$/x$, but of most importance later on is their asymptotic forms for large and small arguments:

$$j_\ell(x) \to x^\ell \qquad \text{and} \qquad y_\ell(x) \to -x^{-\ell-1} \qquad \text{as } x \to 0\,, \tag{3.57}$$

[20] These are solutions provided $U(r)$ falls for large r sufficiently quickly, though $U(r) \propto 1/r$ turns out not to be quick enough. This need not matter in practice for scattering problems since most quantities of interest depend only on the large-r asymptotic form of $j_\ell(kr)$ and $y_\ell(kr)$, which remain valid even when $U(r) \propto 1/r$.

while

$$j_\ell(x) \to \frac{\sin(x-\pi\ell/2)}{x} \qquad \text{and} \qquad y_\ell(x) \to -\frac{\cos(x-\pi\ell/2)}{x} \qquad \text{as } x\to\infty\,. \tag{3.58}$$

It is the large-x version that is of interest when enforcing the boundary condition at large r, and using (3.58) in (3.56) shows that

$$\begin{aligned} R_\ell &\to \frac{A_\ell\,\sin(kr-\pi\ell/2) - B_\ell\,\cos(kr-\pi\ell/2)}{kr} \\ &= C_\ell\,\frac{\sin(kr-\pi\ell/2+\delta_\ell)}{kr}\,, \end{aligned} \tag{3.59}$$

where the second equality makes a conventional change of notation for the integration constants: $A_\ell = C_\ell\cos\delta_\ell$ and $B_\ell = -C_\ell\sin\delta_\ell$.

The above considerations make the angular dependence of ψ at large r more explicit, since

$$\psi(r,\theta) \to \sum_{\ell=0}^{\infty} C_\ell\,\left[\frac{\sin(kr-\pi\ell/2+\delta_\ell)}{kr}\right] P_\ell(\cos\theta) \qquad (\text{large } r)\,. \tag{3.60}$$

To determine $w(\theta,\phi)$ we determine C_ℓ using the boundary condition that this last form must agree with (3.44), and to do this we must also expand[21] the in-coming wave $\psi_i = Ce^{-i(Et-kz)}$ in terms of the $\psi_{\ell\ell_z}$'s. The required expansion is

$$\begin{aligned} e^{ikz} = e^{ikr\cos\theta} &= \sum_{\ell=0}^{\infty} i^\ell(2\ell+1)\,j_\ell(kr)\,P_\ell(\cos\theta) \\ &\to \sum_{\ell=0}^{\infty} i^\ell(2\ell+1)\,\left[\frac{\sin(kr-\pi\ell/2)}{kr}\right] P_\ell(\cos\theta) \qquad (\text{large } r)\,. \end{aligned} \tag{3.61}$$

To fix the C_ℓ's we demand that (3.60) approach the sum of (3.61) and the out-going wave, $w(\theta,\phi)\,e^{ikr}/r$. For this to be possible it must be that all of the *in-coming* waves in (3.60) are equal to those in (3.61) (for each ℓ) once we expand the sine and cosine in terms of $e^{\pm ikr}$. Once this is done we collect the terms in front of the out-going wave in the *difference* between (3.60) and (3.61) to read off w. Equating the coefficients of in-coming waves gives

$$C_\ell = i^\ell(2\ell+1)\,e^{i\delta_\ell}\,, \tag{3.62}$$

and using these for the out-going waves then gives our desired expression for w:

$$w(\theta,\phi) = w(\theta) = \sum_{\ell=0}^{\infty}(2\ell+1)w_\ell(k)\,P_\ell(\cos\theta) \qquad \text{with} \qquad w_\ell(k) = \frac{1}{k}\,e^{i\delta_\ell}\,\sin\delta_\ell\,. \tag{3.63}$$

[21] This must be possible because plane waves are a solution to the free Schrödinger equation, and the series in $\psi_{\ell\ell_z}$ gives the most general solution.

This expansion of $w(\theta)$ as a sum over angular-momentum quantum number is called the *partial-wave* expansion, and δ_ℓ is called the ℓ-th *phase-shift*. In principle everything is determined once the phase-shift is known, and this is found by solving the Schrödinger equation in the presence of $U(r)$ and using this to determine the ratio of integration constants, B_ℓ/A_ℓ, appearing in (3.56). Then we fix δ_ℓ using its definition

$$\tan\delta_\ell = -B_\ell/A_\ell\,. \tag{3.64}$$

When tracking the k-dependence it is worth keeping in mind that δ_ℓ is itself a function of k.

Inserting (3.63) into the cross-section formula, (3.51), then gives a partial-wave decomposition of σ:

$$\begin{aligned}\sigma &= \int \mathrm{d}\Omega \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right) = 2\pi\int_0^\pi |w(\theta)|^2 \sin\theta\,\mathrm{d}\theta \\ &= 4\pi\sum_{\ell=0}^{\infty}(2\ell+1)|w_\ell(k)|^2 = \frac{4\pi}{k^2}\sum_{\ell=0}^{\infty}(2\ell+1)\sin^2\delta_\ell\,,\end{aligned} \tag{3.65}$$

which performs the θ integral using the orthogonality of the Legendre polynomials:

$$\int_0^\pi P_r(\cos\theta)P_s(\cos\theta)\,\sin\theta\,\mathrm{d}\theta = \frac{2}{2r+1}\,\delta_{rs}\,. \tag{3.66}$$

In principle we now have all the information needed to explicitly compute the scattering properties given an interaction potential, $U(r)$.

Notice that using (3.63) to evaluate w in the forward direction — *i.e.* at $\theta = 0$, using $P_\ell(1) = 1$ — gives a result very similar to (3.65):

$$w(\theta=0) = \frac{1}{k}\sum_{\ell=0}^{\infty}(2\ell+1)e^{i\delta_\ell}\sin\delta_\ell\,, \tag{3.67}$$

and so we see — *for any scattering potential* $U(r)$ — that

$$\sigma = \frac{4\pi}{k}\,\mathrm{Im}\;w(\theta=0)\,. \tag{3.68}$$

This very general result is called the *optical theorem* and reflects the unitarity of quantum mechanics. That is, the sum of all of the scattering probabilities must correspond to the depletion of the probability of not scattering, and so continuing on into the forward direction.

3.2.5 Hard-sphere scattering

As our first example consider again scattering from a hard sphere: $U = 0$ for $r > R$ and $U \to \infty$ for $r < R$. In this case the radial solution exterior to the sphere is given by (3.56) and we must impose the boundary condition that $\psi(r = R, \theta, \phi) = 0$ for all values of θ and ϕ. This implies that $R_\ell(r = R) = 0$ for each ℓ and so we read off

$$\tan\delta_\ell = -\frac{B_\ell}{A_\ell} = \frac{j_\ell(kR)}{y_\ell(kR)}\,. \tag{3.69}$$

In principle this is the answer and we can go home.

Notice that the trigonometric functions in j_ℓ and y_ℓ imply this expression predicts an oscillatory structure (superimposed on an overall $1/k^2$ envelope) for scattering cross sections when they are regarded as a function of $k = \sqrt{2mE}$. Physically these oscillations are due to wave-like diffraction of the probability amplitude as it passes the hard sphere.

This diffraction is particularly strong in the low-energy limit, $kR \ll 1$, for which the wavelength of the incident particles is larger than the size of the target. In this limit we use the small-x asymptotic forms for the spherical Bessel functions, (3.57), to find

$$\tan\delta_\ell \simeq \delta_\ell \simeq -(kR)^{2\ell+1} \ll 1\,, \tag{3.70}$$

which shows that only the first few terms of the partial-wave expansion are important at low energies. (This makes the low-energy limit one for which expanding in partial waves is most useful.)

Keeping only the leading power of kR we can drop all but $\ell = 0$ (what is called S-wave scattering), and read off

$$w(\theta) \simeq w_0(k) = -\frac{kR}{k} = -R \qquad \text{and so} \qquad \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq |w_0|^2 = R^2\,. \tag{3.71}$$

This shows that the scattering at low energies is isotropic — as must be $\ell = 0$ scattering, since $P_0(\cos\theta) = 1$ — and that the total cross section goes to the finite low-energy value

$$\sigma \simeq 4\pi R^2\,. \tag{3.72}$$

This is *larger* than the classical result (the geometrical target area) by a factor of 4. It is larger because the diffraction of the incident wave around the target allows the target to influence evolution at distances beyond its geometrical size.

> **Exercise 3.7:** Calculate the differential cross section $\mathrm{d}\sigma/\mathrm{d}\Omega$ for s-wave (*i.e.* $\ell = 0$) quantum scattering from a repulsive delta-function potential $U(\mathbf{r}) = \kappa\,\delta^3(\mathbf{r})$ (with $\kappa > 0$) using the Schrödinger equation. [Hint: you can trade the delta-function potential for the boundary condition $4\pi r^2 \partial_r\psi = 2m\kappa\,\psi$ for infinitesimal (but nonzero) $r = \epsilon$.]

3.2.6 Coulomb scattering

The Schrödinger equation for a Coulomb potential, $U = \kappa_c/r$, can be exactly solved in closed form in spherical coordinates, much as described above. In this case the radial wave-functions are given in terms of confluent hypergeometric functions and the partial-wave phase shifts work out to be (see any undergraduate Quantum text)

$$e^{2i\delta_\ell} = \frac{\Gamma(\ell+1-i\mu)}{\Gamma(\ell+1+i\mu)}\,, \tag{3.73}$$

where $\mu := m\kappa_c/k = \kappa_c/v_i$ and $\Gamma(z)$ is Euler's gamma function. $w(\theta)$ is found by performing the sum over ℓ (or by solving directly for it in the Schrödinger equation using parabolic coordinates rather than spherical polar coordinates) and gives

$$w(\theta) = \frac{\mu\, e^{i\eta}}{2k\sin^2(\theta/2)}\,, \tag{3.74}$$

where the phase is $\eta := -\mu\ln[\sin^2(\theta/2)] + \pi + 2\arg\Gamma(1+i\mu)$, where 'arg' means 'the phase of' a complex number.

The corresponding differential cross section becomes

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = |w(\theta)|^2 = \left(\frac{\kappa_c}{2mv_i^2}\right)^2 \csc^4\frac{\theta}{2}\,, \tag{3.75}$$

in perfect agreement with the classical result, (3.22). (It was apparently a great source of pride for Rutherford that his formula was one of the few that survived the advent of quantum mechanics.)

> **Exercise 3.8:** Solve the 3-dimensional time-independent Schrödinger equation for a particle of mass m interacting with the repulsive inverse-square potential, $U(r) = \kappa/r^2$, with $\kappa > 0$. Separate variables in spherical polar coordinates, $\psi(r,\theta,\phi) = R(r)Y_{\ell\ell_z}(\theta,\phi)$ and derive the radial ordinary differential equation satisfied by $R(r)$. In what ways does the result resemble the radial equation in the absence of a potential? Based on this, what are the general solutions to the radial equation for the inverse-square potential? For $\ell = 0$ what is the solution that ensures $\psi(r,\theta,\phi)$ is nonsingular at $r = 0$ and goes over to the expected answer when $\kappa \to 0$? What is the s-wave differential scattering cross section, $\mathrm{d}\sigma/\mathrm{d}\Omega$, for $\ell = 0$ and small $m\kappa$?

3.2.7 An attractive square well

Consider next a finite square well, with $U = 0$ for $r > R$ and $U = -U_0$ for $r < R$. Besides being solvable, as we shall see this is a poor man's model of nuclear forces: attractive but with finite range. When solving the Schrödinger equation for this potential the heavy lifting comes when we solve the radial equation, (3.52), which simplifies a bit if we define $u_\ell(r)$ by $R_\ell(r) = u_\ell(r)/r$ to become

$$\frac{\mathrm{d}^2u_\ell}{\mathrm{d}r^2} + \left[2mE - \frac{\ell(\ell+1)}{r^2} - 2mU(r)\right]u_\ell = 0\,. \tag{3.76}$$

The case $\ell = 0$ (which we've seen should dominate in any case at low energies or for short ranges — *i.e.* when $kR \ll 1$) is particularly simple, with general solution for $u(r) := u_0(r)$ given by

$$u(r) := A\cos(kr) + B\sin(kr) \qquad \text{with} \qquad k = \sqrt{2m(E-U)} \tag{3.77}$$

and A and B integration constants.

There are several cases to consider. First, since U differs inside and outside the well we take solutions of this form separately in these two regions, after which we must demand continuity of both $u(r)$ and $u'(r)$ across $r = R$. The kind of solutions obtained outside the well depend on whether or not E is positive or not (and so on whether k is real or imaginary). Although our main focus is scattering ($E > 0$) we consider both cases in turn in order to connect with an earlier result.

Bound states: $-U_0 < E < 0$

In this case we write $E = -E_B$ with $U_0 > E_B > 0$ and $k = i\kappa$ for $r > R$ so that the exterior solution becomes

$$u_{\rm out}(r) = A_{\rm out}\, e^{+\kappa r} + B_{\rm out}\, e^{-\kappa r} \qquad \text{where} \qquad \kappa = \sqrt{2mE_B}\,. \tag{3.78}$$

and so normalizability requires we take $A_{\rm out} = 0$. For $r < R$ we instead have (3.77) with $k = \sqrt{2m(E-U)} = \sqrt{2m(U_0 - E_B)}$ real. In order for $R = u/r$ to remain finite at $r = 0$ we take $A_{\rm in} = 0$, and so $u_{\rm in} = B_{\rm in}\sin(kr)$. Continuity of u and u' at $r = R$ then implies

$$B_{\rm out} e^{-\kappa R} = B_{\rm in}\sin(kR) \qquad \text{and} \qquad -B_{\rm out}\kappa e^{-\kappa R} = B_{\rm in} k\cos(kR)\,, \tag{3.79}$$

which we can solve for $B_{\rm in}/B_{\rm out}$, but not $B_{\rm in}$ and $B_{\rm out}$ separately (which we instead determine from the normalization condition). They cannot both be determined because the ratio of these equations does not depend on them,

$$\kappa R = -(kR)\cot(kR)\,, \tag{3.80}$$

and this equation instead imposes a quantization condition on κ (and so also on E). This is a quantization condition and not just a relationship between k and κ because their definitions — $\kappa^2 = 2mE_B$ and $k^2 = 2m(U_0 - E_B)$ — imply that they are both already determined by the one quantity E_B. Equivalently, kR and κR must satisfy both (3.80) and

$$(kR)^2 + (\kappa R)^2 = 2mU_0R^2\,. \tag{3.81}$$

Although (3.80) is a transcendental equation for E_B, its implications can be seen graphically by plotting both it and (3.81) and asking when the resulting curves intersect, as in Figure 20. This shows that the number of intersections (and so the number of bound states) depends on the value of $2mU_0R^2/\hbar^2$ (where the factors of $\hbar$ are put back as required by dimensional analysis). The prediction is

$$\begin{aligned}
\frac{\sqrt{2mU_0R^2}}{\hbar} &< \frac{\pi}{2} && \text{(no bound state)} \\
\frac{\pi}{2} < \frac{\sqrt{2mU_0R^2}}{\hbar} &< \frac{3\pi}{2} && \text{(one bound state)} \\
\frac{3\pi}{2} < \frac{\sqrt{2mU_0R^2}}{\hbar} &< \frac{5\pi}{2} && \text{(two bound states)}\,,
\end{aligned} \tag{3.82}$$

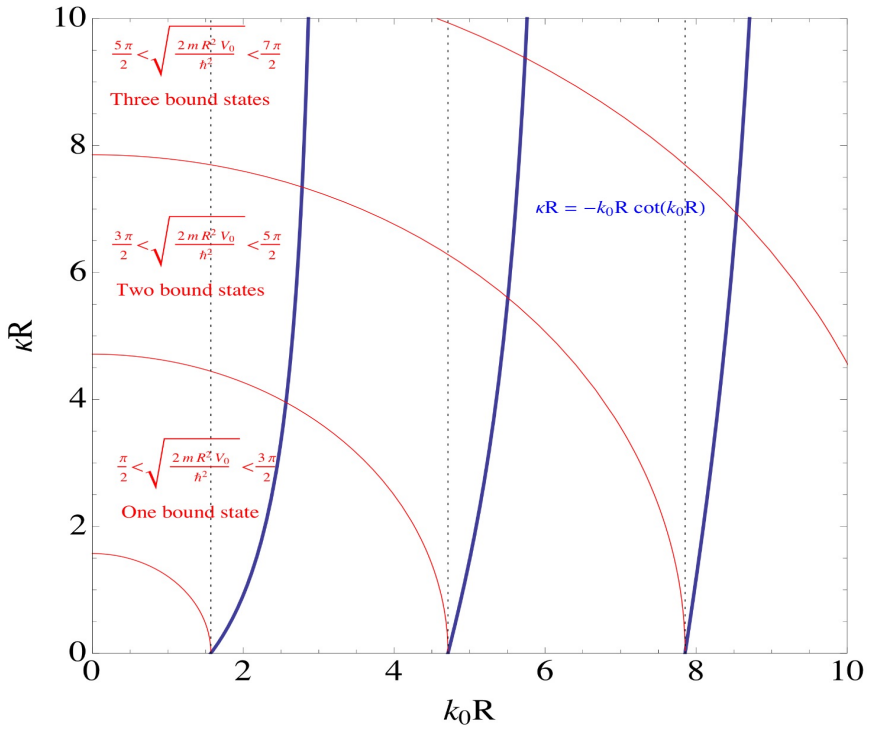


Figure 20. A graphical determination of the bound-state energy quantization condition. Allowed energies correspond to when the circles and tangents intersect. (Figure source: Itay Yavin.)

and so on. That is, if the depth, U_0, of the well is too shallow then it cannot compensate for the zero-point energy associated with localizing the particle within a radius R: $2mE_{\rm kin} \sim (\hbar/R)^2$, imposed by the uncertainty principle. And the deeper the well is the more wavelengths can be fit into it without costing so much kinetic energy that the particle escapes.

For future use notice also that $kR = \left(n + \frac{1}{2}\right)\pi$ satisfies $\cot(kR) = 0$ for any integer n and so corresponds to a solution for which $\kappa = E_B = 0$.

Scattering states: $E > 0$

Now we turn to the scattering states, so have inside and outside solutions of the following form

$$u_{\rm in}(r) = B_{\rm in}\,\sin(k_{\rm in} r) \quad \text{for } r < R \qquad \text{and} \qquad u_{\rm out}(r) = C_{\rm out}\,\sin(kr + \delta_0) \quad \text{for } r > R\,, \tag{3.83}$$

where $k_{\rm in} = \sqrt{2m(E + U_0)}$ and $k = \sqrt{2mE}$. We use the finiteness of $R(r)$ at $r = 0$ to eliminate the integration constant $A_{\rm in}$, as above, and trade the constants $A_{\rm out}$ and $B_{\rm out}$ of the external solution for $C_{\rm out}$ and δ_0, since this is the form of the outgoing wave we seek at large r.

Continuity of u and u' across $r = R$ then gives

$$\begin{aligned} C_{\rm out}\sin(kR+\delta_0) &= B_{\rm in}\sin(k_{\rm in}R) \\ \text{and} \qquad kC_{\rm out}\cos(kR+\delta_0) &= k_{\rm in}B_{\rm in}\cos(k_{\rm in}R)\,. \end{aligned} \tag{3.84}$$

One combination of these gives $C_{\rm out}/B_{\rm in}$, while their ratio imposes a condition independent of these that can be used to fix δ_0 as a function of E:

$$\tan(kR+\delta_0) = \frac{k}{k_{\rm in}}\,\tan(k_{\rm in}R)\,. \tag{3.85}$$

It is at low energies that we expect this $\ell = 0$ analysis to dominate, and if $E \ll U_0$ then $k/k_{\rm in} \simeq \sqrt{E/U_0} \ll 1$ and so (assuming $\tan(k_{\rm in}R)$ is not too large — a choice we later relax) then $kR + \delta_0$ is small. Then the solution for δ_0 is

$$kR+\delta_0 \simeq \frac{k}{k_{\rm in}}\,\tan(k_{\rm in}R) \qquad \text{and so} \qquad \delta_0 \simeq kR\left[\frac{\tan(k_{\rm in}R)}{k_{\rm in}R} - 1\right]. \tag{3.86}$$

Because $\ell = 0$ the scattering is isotropic, and the total cross section is

$$\sigma \simeq \sigma_0 = \frac{4\pi}{k^2}\,\sin^2\delta_0 \simeq \frac{4\pi\delta_0^2}{k^2} \simeq 4\pi R^2\left[\frac{\tan(k_{\rm in}R)}{k_{\rm in}R} - 1\right]^2, \tag{3.87}$$

where $k_{\rm in} = \sqrt{2m(E+U_0)} \simeq \sqrt{2mU_0}$. At low energies the cross section goes to a constant, whose value is roughly set by the range, R, of the potential, up to order-unity diffraction effects. (Recall we assumed $\tan(k_{\rm in}R)$ not to be large in its derivation.)

3.2.8 Resonance

The exception to the constancy of σ in the low-energy limit is when the $\tan(k_{\rm in}R)$ factor is *not* order unity in the above argument. This happens whenever $E = E_\star$ is such that $k_{\rm in}R$ is very close to $\left(n+\frac{1}{2}\right)\pi$, which we saw above is also the criterion for there to be a bound state very close to $E = 0$. In this case we know that $\tan(k_{\rm in}R)$ can become very large despite the smallness of $k/k_{\rm in}$ in the low-energy limit. Consequently the left side of (3.85) must also become large, and because we are in the low-energy limit where $kR \ll 1$, it follows that $\delta_0(E)$ must go to $\frac{\pi}{2}$ at this energy in order for the tangent function to blow up.

This implies the cross section acquires an energy dependence near these points, rather than being approximately constant as it is otherwise. More generally, Taylor expanding $\cot\delta_\ell$ near $E = E_\star$ we then find:

$$\cot\delta_\ell(E) \simeq -\frac{2}{\Gamma}\,(E - E_\star) + \cdots, \tag{3.88}$$

where $-2/\Gamma$ proves to be a convenient way to write the Taylor coefficient. Using this expansion in the scattering amplitude then gives

$$w_\ell = \frac{1}{k}\,e^{i\delta_\ell}\sin\delta_\ell = \frac{1}{k}(\cos\delta_\ell + i\sin\delta_\ell)\sin\delta_\ell = \frac{1}{k}\left[\frac{\sin\delta_\ell}{\cos\delta_\ell - i\sin\delta_\ell}\right] = \frac{1}{k}\left[\frac{1}{\cot\delta_\ell - i}\right], \tag{3.89}$$

and so

$$w_\ell(E) \simeq -\frac{1}{k}\left[\frac{\Gamma/2}{E - E_\star + i\Gamma/2}\right] . \tag{3.90}$$

This is a form we have seen before. The total cross section near $E = E_\star$ that follows from this has the Breit-Wigner form,

$$\sigma \simeq \frac{4\pi}{k^2}(2\ell+1)\left[\frac{(\Gamma/2)^2}{(E-E_\star)^2 + (\Gamma/2)^2}\right] , \tag{3.91}$$

that we saw is characteristic of a resonance. It indicates an intermediate state with a decay rate Γ that is being produced and then decays into the observed final state. In the present example the corresponding intermediate state is the bound state found above for the potential near $E = 0$.

3.2.9 Scattering length and effective range

The low-energy expansion of the cross section is also often written in terms of an expansion of $\cot\delta_0$ in powers of k (that is, expanding about $E = 0$):

$$k\cot\delta_0 \simeq -\frac{1}{a_0} + \frac{r_0 k^2}{2} + \cdots , \tag{3.92}$$

and we build in that the cross section approaches a finite limit as $k \to 0$.

The parameter a_0 is called the *scattering length* and r_0 is called the *effective range*, and the two of them parameterize the low-energy limit of the cross section. The definition of the scattering length in particular is chosen so that the low-energy cross section becomes

$$\sigma \simeq \sigma_0 = \left(\frac{4\pi}{k^2}\right)\frac{1}{1+\cot^2\delta_0} \quad \to 4\pi a_0^2 \text{ as } k \to 0 , \tag{3.93}$$

and so a_0 appears as does the radius of a hard sphere.

For nuclear interactions a typical value for a_0 is the nuclear length scale, $a_0 \simeq 1$ fm. Such a value leads to the expectation that nuclear scattering cross sections (for *e.g.* neutrons, which do not experience long-range Coulomb forces) should asymptote to constants for $k \ll 1/a_0$ (in the absence of very low-energy resonances). Furthermore, the size expected for this constant cross section is of order

$$\sigma \simeq 4\pi a_0^2 \sim 12 \text{ fm}^2 = 1.2 \times 10^{-25} \text{ cm}^2 = 0.12 \text{ b} , \tag{3.94}$$

showing again why the barn is a useful unit in subatomic physics. Scattering cross sections for neutron scattering from nuclei of several elements are plotted as solid lines in Fig. 21, and indeed become energy independent[22] with values close to 1 b for a wide range of energies below 1 MeV.

[22] Close inspection of Fig. 21 shows that elastic cross sections start to become proportional to $1/k$ at neutron energies below 0.01 eV. This marks a transition to an energy range for which the electromagnetic interaction between the neutron magnetic moment and the nucleus' charge starts to become larger than scattering due to nuclear interactions.

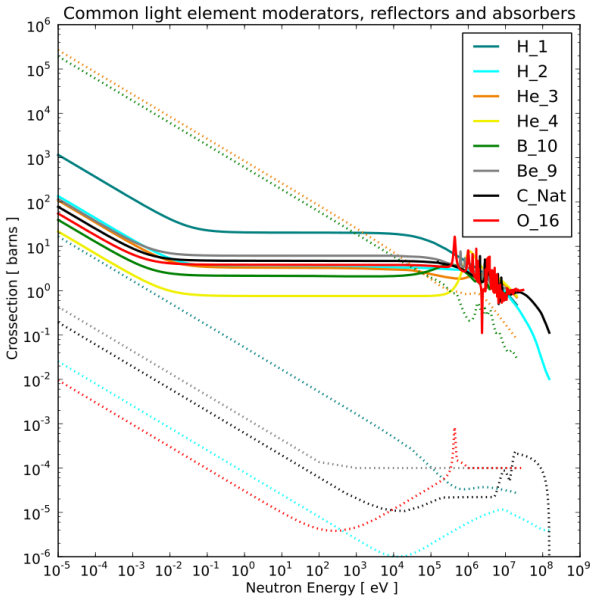


Figure 21. A plot of neutron scattering and absorption cross sections *vs* neutron energy for a variety of elements found in reactor materials. Solid (dotted) lines correspond to scattering (absorption) cross sections, while colours distinguish results for different elements. Note the comparatively high absorption cross sections for ^{3}He and ^{10}B. Scattering cross sections remain roughly constant – in agreement with eq. (3.93) – for the low-energy range 10^{-2} eV $< E < 10^6$ eV. For the same energy range absorption cross sections instead vary like $\sigma \propto E^{-1/2} \propto 1/k$. (Figure source: Wikipedia `https://en.wikipedia.org/wiki/Neutron_cross_section`).

3.2.10 Low-energy absorption cross sections

The above discussion seems to make it inevitable that (in the absence of resonances) low-energy nuclear cross sections become energy-independent, but Fig. 21 also shows that this is *not* true for neutron absorption cross sections – which describe reactions where the incoming neutron is absorbed by the nucleus as it transitions to a new nuclear state. For instance, an example of an absorptive reaction for neutrons might be the reaction

$$n + {}^3\mathrm{He} \to {}^4\mathrm{He} + \gamma\,, \tag{3.95}$$

which is often written in the short form ^{3}He (n,γ) ^{4}He. Fig. 21 reveals these types of reactions to vary with initial neutron energy like $\sigma_{\rm abs} \propto E^{-1/2} \propto k^{-1}$ for small k – even in the range for which scattering cross sections remain roughly constant. Why does the above argument fail for absorption?

There are two main assumptions in the arguments leading to (3.93). One is the absence of resonances; or more generally access at low energies to intermediate states that can allow nontrivial momentum dependence. It is this loophole that Coulomb scattering threads, for

example, with its low-energy dependence $\mathrm{d}\sigma/\mathrm{d}\Omega \propto 1/k^2$ as found in eq. (3.75). In this case it is virtual-photon exchange that provides the intermediate channel that precludes having constant cross section at low energy.

The second key assumption used in (3.93) is that the scattering is elastic. The role of elasticity can be seen from kinematic expressions like (2.53) or (2.54) for scattering, which give results proportional to k'/k where k' is a final-state momentum, while k is the momentum of the initial state. As the derivation of these equations makes clear these factors do not depend on the details of any matrix elements; the factor of k' comes from the final-state phase space integration while the $1/k$ comes from the division by $v_{\rm rel}$ contained when dividing out the flux to obtain the cross section from the reaction rate. What is important about elastic scattering is that $k' = k$ and so these factors cancel. This need not be so for inelastic processes, for which $k \to 0$ need not also imply that k' vanishes. Absorptive scattering exploits this loophole to give $\sigma \propto 1/k$, as we now show.

A complete description of absorptive scattering requires tracking the new degrees of freedom whose production carries off the energy and momentum deposited by the absorbed particle. The discussion is a bit simpler, though, if one asks only what are called 'inclusive' questions, that measure only the depletion of probability from the absorbed particle's initial state, without also asking what the rest of the system does. When restricted only to the initial absorbed-particle sector, the evolution is not unitary in that conservation of probability breaks down: the sum over the probabilities to scatter into all possible final momenta plus the probability of not scattering at all for this particle gives a result smaller than unity (because this does not count the final-state channels produced by absorption).

In particular, unitarity is assumed when deriving the scattering amplitude for each partial wave, leading to the expression (3.63):

$$w_\ell = \frac{1}{k}\, e^{i\delta_\ell} \sin\delta_\ell \quad \text{and so} \quad 0 \le |w_\ell| \le \frac{1}{k}\,, \tag{3.96}$$

where the inequalities follow because the phase angle δ_ℓ is real. It is perhaps not a surprise, then, that a way to parameterize absorptive cross sections is to entertain situations where the scattering phase acquires a nonzero (and positive) imaginary part. It is conventional then to use the symbol γ_ℓ for the complete complex scattering 'phase' and reserve δ_ℓ only for its real part, so $\gamma_\ell = \delta_\ell + i\eta_\ell$ with $\eta_\ell \ge 0$ and

$$e^{i\gamma_\ell} = e^{-\eta_\ell}\, e^{i\delta_\ell}\,. \tag{3.97}$$

Whereas eq. (3.96) implies w_ℓ satisfies $0 \le |w_\ell| \le 1/k$ holds for elastic scattering, for absorptive scattering instead one has

$$w_\ell = \frac{1}{k}\, e^{i\gamma_\ell} \sin\gamma_\ell = \frac{i}{2k}\Bigl(1 - e^{2i\gamma_\ell}\Bigr) = \frac{i}{2k}\Bigl(1 - e^{-2\eta_\ell}\, e^{2i\delta_\ell}\Bigr)\,, \tag{3.98}$$

which now satisfies the stricter inequality

$$0 \le |w_\ell| \le \frac{1}{2k}\Big(1 + e^{-2\eta_\ell}\Big) < \frac{1}{k}\,, \tag{3.99}$$

reflecting (for $\eta_\ell > 0$) a loss of probability compared to pure elastic scattering.

So far so good, but how does this change the low-energy limit? As usual at low energies it is the $\ell = 0$ mode that dominates, and (as before) for small k the scattering amplitude for this is suppressed linearly in k:

$$\gamma_0 \simeq (a_0 + ib_0)\,k + \mathcal{O}(k^2) \quad \text{and so} \quad e^{i\gamma_0} \simeq 1 + i(a_0 + ib_0)\,k + \mathcal{O}(k^2)\,. \tag{3.100}$$

The elastic-scattering cross section therefore asymptotes as before to a constant:

$$\sigma_{\rm el} \simeq 4\pi|w_0|^2 = \frac{\pi}{k^2}\Big|1 - e^{2i\gamma_0}\Big|^2 \simeq 4\pi|a_0 + ib_0|^2\,. \tag{3.101}$$

But the absorption cross section, by contrast, is

$$\sigma_{\rm ab} = \frac{\pi}{k^2}\sum_{\ell=0}^{\infty}(2\ell+1)\left(1 - e^{-4\eta_\ell}\right), \tag{3.102}$$

and so at low energies

$$\sigma_{\rm ab} \simeq \frac{\pi}{k^2}\Big(1 - e^{-4\eta_0}\Big) \simeq \frac{4\pi\, b_0}{k}\,, \tag{3.103}$$

which varies as $1/k$ for small k, as seen in Fig. 21. So although both elastic and absorptive parts of the scattering amplitude start off like $\gamma_0 \propto k$ for small k, the difference between the low-energy dependence of the elastic and absorptive cross sections schematically arises because the elastic cross section is proportional to $(1/k)^2|e^{i\gamma_0} - 1|^2 \sim |\gamma_0|^2/k^2$ while the absorptive cross section is proportional to $(1/k^2)[|e^{i\gamma_0}|^2 - 1] \sim \mathrm{Im}\,\gamma_0/k^2$.

This growth for small k was discovered in the 1930s by Hans Bethe, who before this discovery had been worried that quantum mechanics might not be able to produce low-energy neutron absorption cross sections that were large enough to agree with observations.

3.3 Perturbation theory and the Born approximation

We have seen so far that given an interaction potential, $U(r)$, we can compute a differential scattering cross section, $\mathrm{d}\sigma/\mathrm{d}\Omega$, and so measurements of the dependence of the scattering rate on energy and angle provide information about the form of the interaction responsible for the scattering. However the story so far has the drawback that the connection between the cross section and interaction is fairly indirect and not explicit; and potentially challenging to compute for real examples. This section develops a perturbative solution for scattering applicable to weak interactions, along the lines of the impulse approximation described earlier. When it applies it provides a very direct connection between scattering amplitudes and interaction potentials.

3.3.1 Green's functions

We seek solutions to

$$-\frac{1}{2m}\nabla^2\psi + U(\mathbf{r})\,\psi = E\psi\,, \tag{3.104}$$

perturbatively in powers of U. That is, formally we write $U = \epsilon U$ and take $\psi = \psi_0 + \epsilon\psi_1 + \epsilon^2\psi_2 + \cdots$ and substitute this into (3.104). Demanding the solution to hold for all ϵ allows us to separately set to zero the coefficient of each power, leading to the sequence of equations

$$\begin{aligned}\nabla^2\psi_0 + 2mE\,\psi_0 &= 0\\ \nabla^2\psi_1 + 2mE\,\psi_1 &= 2mU(\mathbf{r})\,\psi_0\\ \nabla^2\psi_2 + 2mE\,\psi_2 &= 2mU(\mathbf{r})\,\psi_1\,,\end{aligned} \tag{3.105}$$

and so on. For scattering problems we imagine solving the first of these with the incoming plane wave, $\psi_0 = e^{ikz}$, with $k = \sqrt{2mE}$ as usual (for positive E). Then we regard the next equation as to be solved for ψ_1 given ψ_0; the third to be solved for ψ_2 given ψ_1 and so on.

We focus here on the 2nd equation for ψ_1. This has the structure of $\hat{O}\,\psi_1 = J$ where $\hat{O}$ is the differential operator and J is the right-hand side. If this had been a matrix equation, with $\hat{O}$ a matrix and ψ_1 and J vectors, then the solution would immediately have been $\psi_1 = \hat{O}^{-1}J$. We seek the 'inverse matrix' (or *Green's function*) for the differential operator $\hat{O}$.

To construct this we start with the equation in the form

$$(\nabla^2 + \mu^2)\psi_1 = J\,, \tag{3.106}$$

where $\mu = \sqrt{2mE}$ and $J = 2mU\psi_0$. It is useful to Fourier transform the equation and write

$$\psi_i(\mathbf{r}) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,\tilde{\psi}_i(\mathbf{k})\,e^{i\mathbf{k}\cdot\mathbf{r}} \qquad \text{with inverse} \qquad \tilde{\psi}_i(\mathbf{k}) = \int \mathrm{d}^3\mathbf{r}\,\psi_i(\mathbf{r})\,e^{-i\mathbf{k}\cdot\mathbf{r}}\,, \tag{3.107}$$

and similarly for $J(\mathbf{r})$ in terms of $\tilde{J}(\mathbf{k})$. Proving these are inverse transformations of one another (and an explanation of the factors of 2π) uses the orthogonality of plane waves,

$$\int \mathrm{d}^3\mathbf{r}\;e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} = (2\pi)^3\,\delta^3(\mathbf{k}-\mathbf{k}')\,. \tag{3.108}$$

The utility of this transform is that (3.106) becomes an algebraic equation,

$$(-\mathbf{k}^2 + \mu^2)\tilde{\psi} = \tilde{J}\,, \tag{3.109}$$

that is easy to solve:

$$\tilde{\psi}(\mathbf{k}) = \frac{\tilde{J}(\mathbf{k})}{-\mathbf{k}^2 + \mu^2}\,. \tag{3.110}$$

As is easy to check, transforming back to $\psi(\mathbf{r})$ and $J(\mathbf{r})$ this last equation is equivalent to the convolution

$$\psi(\mathbf{r}) = \int \mathrm{d}^3\mathbf{x}\;G(\mathbf{r}-\mathbf{x})\,J(\mathbf{x})\,, \tag{3.111}$$

where the Green's function, G, is given by

$$G(\mathbf{r}-\mathbf{x}) = -\int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,\frac{e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{x})}}{\mathbf{k}^2-\mu^2+i\varepsilon}\,. \tag{3.112}$$

Here ε is a small positive quantity that is taken to zero at the end of the calculation, whose role is to clearly specify the integration range of a later integral (that is otherwise ill-defined because of the singularity of the integrand when $\mathbf{k}^2=\mu^2$).

The function $G(\mathbf{r}-\mathbf{x})$ is the 'inverse matrix' for the differential operator, $\hat{O}=\nabla^2+\mu^2$, as sought, in the following sense:

$$\hat{O}\,G(\mathbf{r}-\mathbf{x}) = (\nabla^2+\mu^2)G(\mathbf{r}-\mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{x})} = \delta^3(\mathbf{r}-\mathbf{x})\,, \tag{3.113}$$

and using this in (3.111) shows that (3.111) indeed solves the differential equation (3.106).

$G(\mathbf{r}-\mathbf{x})$ can be written as a closed-form function of position by doing the integrals over $\mathbf{k}$ explicitly using polar coordinates for $\mathbf{k}=\{k,\vartheta,\varphi\}$ with the z-axis defined in the direction defined by $\mathbf{r}-\mathbf{x}$. This then gives $(\mathbf{r}-\mathbf{x})\cdot\mathbf{k} = k|\mathbf{r}-\mathbf{x}|\cos\vartheta$ and $\mathrm{d}^3k = \mathrm{d}k_x\mathrm{d}k_y\mathrm{d}k_z = k^2\sin\vartheta\,\mathrm{d}k\mathrm{d}\vartheta\mathrm{d}\varphi$, and since the integrand does not depend on φ its integral corresponds to multiplying by 2π, leaving:

$$\begin{aligned} G(\mathbf{r}-\mathbf{x}) &= -\frac{1}{4\pi^2}\int_0^\infty \mathrm{d}k\,\frac{k^2}{k^2-\mu^2+i\varepsilon}\int_{-1}^{1}\mathrm{d}\cos\vartheta\;e^{ik|\mathbf{r}-\mathbf{x}|\cos\vartheta} \\ &= -\frac{1}{2\pi^2|\mathbf{r}-\mathbf{x}|}\int_0^\infty \mathrm{d}k\,\frac{k}{k^2-\mu^2+i\varepsilon}\,\sin(k|\mathbf{r}-\mathbf{x}|)\,. \end{aligned} \tag{3.114}$$

It is this last integral that the ε is designed to make well-defined by shifting the pole in the integrand slightly off the real axis. The result can be evaluated by contour integration to give

$$G(\mathbf{r}-\mathbf{x}) = -\frac{e^{i\mu|\mathbf{r}-\mathbf{x}|}}{4\pi|\mathbf{r}-\mathbf{x}|}\,. \tag{3.115}$$

3.3.2 The perturbative expansion

We are now in a position to express the solution, $\psi(\mathbf{r})$, of our original equation, (3.104), as a series in powers of U. This amounts to solving the tower of equations, (3.105), for the corrections ψ_1, ψ_2 and so on, which can be done using the substitutions $\mu\to k=\sqrt{2mE}$ and $J(\mathbf{x}) = 2mU(\mathbf{x})\psi_0(\mathbf{x})$ in (3.106), whose solutions, (3.111), we have just constructed. This leads to the explicit form

$$\psi_1(\mathbf{r}) = -\frac{m}{2\pi}\int \mathrm{d}^3\mathbf{x}\,\frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}\,U(\mathbf{x})\,\psi_0(\mathbf{x})\,, \tag{3.116}$$

and

$$\psi_2(\mathbf{r}) = -\frac{m}{2\pi}\int \mathrm{d}^3\mathbf{x}\,\frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}\,U(\mathbf{x})\,\psi_1(\mathbf{x})\,, \tag{3.117}$$

and so on.

Assembling these gives the final series form for the scattering state ψ:

$$\begin{aligned}\psi(\mathbf{r}) &= \psi_0(\mathbf{r}) + \psi_1(\mathbf{r}) + \psi_2(\mathbf{r}) + \cdots \\ &= \psi_0(\mathbf{r}) - \frac{m}{2\pi}\int \mathrm{d}^3\mathbf{x}\, \frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}\, U(\mathbf{x})\, \psi_0(\mathbf{x}) \\ &\qquad + \left(\frac{m}{2\pi}\right)^2 \int \mathrm{d}^3\mathbf{x}\, \frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}\, U(\mathbf{x}) \int \mathrm{d}^3\mathbf{y}\, \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}\, U(\mathbf{y})\, \psi_0(\mathbf{y}) + \cdots .\end{aligned} \tag{3.118}$$

The series is often written graphically as in Figure 22. Here each node is labelled by a position, say $\mathbf{x}_i$, and represents a factor of $-(m/2\pi)U(\mathbf{x}_i)$ and each internal line connects two nodes, say $\mathbf{x}_i$ and $\mathbf{x}_j$, and represents a factor of $G(\mathbf{x}_i - \mathbf{x}_j)$. The graph is evaluated by assembling all of these factors and integrating over the positions of each node, and when this is done builds up the expression (3.118). As we will see, a similar kind of graphical expansion of a series solution to scattering problems arises within quantum field theory, in which case the diagrams are called *Feynman diagrams*.

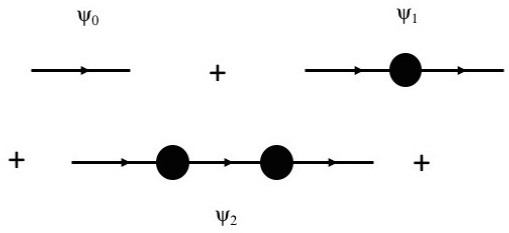


Figure 22. A graphical representation of the perturbative expansion.

3.3.3 The Born approximation

We now apply the above series solution to the scattering problem, and in so doing generate a perturbative *Born expansion* for the scattering state. To this end we start with the zeroth-order (free) solution describing the incoming wave: $\psi_0(\mathbf{x}) = e^{ikz} = e^{i\mathbf{k}_i\cdot\mathbf{x}}$ where $\mathbf{k}_i$ is the initial momentum. The leading correction to this wave (at linear order in U) then is given by (3.116), and so

$$\psi_1(\mathbf{r}) = -\frac{m}{2\pi}\int \mathrm{d}^3\mathbf{x}\, \frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}\, U(\mathbf{x})\, e^{i\mathbf{k}_i\cdot\mathbf{x}}\,. \tag{3.119}$$

For scattering our interest is specifically in the form of this solution at asymptotically large distances, where $r = |\mathbf{r}| \gg |\mathbf{x}|$, where we drop all terms in ψ_1 that fall off faster than $1/r$. Using

$$|\mathbf{r}-\mathbf{x}| = \sqrt{r^2 - 2\mathbf{r}\cdot\mathbf{x} + |\mathbf{x}|^2} \simeq r\left(1 - \frac{\mathbf{r}\cdot\mathbf{x}}{r^2} + \cdots\right) = r\left(1 - \frac{\mathbf{e}_r\cdot\mathbf{x}}{r} + \cdots\right), \tag{3.120}$$

where $\mathbf{e}_r := \mathbf{r}/r$ is the unit vector in the radial direction. The $\mathbf{r}$-dependence of the integrand can be approximated by

$$\frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|} \simeq \left(\frac{e^{ikr}}{r}\right)\, e^{-ik\,\mathbf{e}_r\cdot\mathbf{x}} = \left(\frac{e^{ikr}}{r}\right)\, e^{-i\mathbf{k}_f\cdot\mathbf{x}} \tag{3.121}$$

We find the following expression for the $1/r$ term in the far-field part of $\psi_1(\mathbf{r})$:

$$\begin{aligned}\psi_1(\mathbf{r}) &= -\frac{m}{2\pi}\int \mathrm{d}^3\mathbf{x}\,\frac{e^{ik|\mathbf{r}-\mathbf{x}|}}{|\mathbf{r}-\mathbf{x}|}\,U(\mathbf{x})\,e^{i\mathbf{k}_i\cdot\mathbf{x}} \\ &\simeq -\frac{m}{2\pi}\left(\frac{e^{ikr}}{r}\right)\int \mathrm{d}^3\mathbf{x}\,U(\mathbf{x})\,e^{i(\mathbf{k}_i-\mathbf{k}_f)\cdot\mathbf{x}}\,.\end{aligned} \tag{3.122}$$

Comparing this with $w(\theta)\,e^{ikr}/r$ gives the leading Born approximation for the scattering amplitude

$$\boxed{w(\theta) \simeq -\frac{m}{2\pi}\int \mathrm{d}^3\mathbf{x}\,U(\mathbf{x})\,e^{-i\mathbf{q}\cdot\mathbf{x}} = -\frac{m}{2\pi}\,\tilde{U}(\mathbf{q})}\,, \tag{3.123}$$

where the momentum transfer, $\mathbf{q} := \mathbf{k}_f - \mathbf{k}_i$, has magnitude

$$q = |\mathbf{k}_f - \mathbf{k}_i| = \sqrt{2k^2(1-\cos\theta)} = 2k\,\sin\frac{\theta}{2}\,, \tag{3.124}$$

when expressed in terms of k and the scattering angle: $\mathbf{k}_f\cdot\mathbf{k}_i = k^2\cos\theta$. In the special case where $U(\mathbf{x})$ depends only on $|\mathbf{x}|$ rather than the direction of $\mathbf{x}$ the angular integrals can be done explicitly to give

$$w(\theta) = -\frac{2m}{q}\int_0^\infty \mathrm{d}r\; rU(r)\,\sin(qr)\,. \tag{3.125}$$

We see that the leading contribution to $w(\theta)$ is simply the Fourier transform of the potential evaluated at the momentum transfer of the collision. The leading part of the differential scattering cross section therefore is

$$\boxed{\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq \frac{m^2}{4\pi^2}\left|\tilde{U}(\mathbf{q})\right|^2}\,. \tag{3.126}$$

Exercise 3.9: Calculate the differential cross section $\mathrm{d}\sigma/\mathrm{d}\Omega$ for quantum scattering from a repulsive delta-function potential $U(\mathbf{r}) = \kappa\,\delta^3(\mathbf{r})$ (with $\kappa > 0$) using the Born approximation. If you have evaluated the exact result for this type of scattering – calculated in Exercise 3.7 – how does your result compare?

3.3.4 The Yukawa (or screened) potential

An important example to which the above story can be applied is the case of a *Yukawa* potential,

$$U(r) = \frac{\kappa_c\,e^{-r/a}}{r}\,, \tag{3.127}$$

where a is called the range of the potential. This potential turns out to describe part of the nuclear force, where the range, a, is of the order of a fm. It also arises when a nuclear charge is screened by electrons, in which case a is more of order the Bohr radius (and so of the order of Angstroms). In the limit $a \to \infty$ the Yukawa potential reduces to the Coulomb potential.

Inserting this into (3.125) allows the integral to be performed, giving

$$w(\theta) = -\frac{2m\kappa_c}{q}\int_0^\infty \mathrm{d}r\; e^{-r/a}\,\sin(qr) = -\frac{2m\kappa_c\, a^2}{1+(qa)^2}\,. \tag{3.128}$$

In the limit $a \to \infty$ this becomes $w \to -2m\kappa_c/q^2 = -(m\kappa_c/2k^2)\csc^2(\theta/2)$ in eerie agreement[23] with the Coulomb result.

Conversely, in the limit where the force is very short ranged compared with the momentum transfer — *i.e.* when $qa \to 0$ — the amplitude instead goes to a q-independent constant $w \to -2m\kappa_c a^2$ whose size depends directly on the range of the interaction. In this limit the differential cross section is isotropic, $\mathrm{d}\sigma/\mathrm{d}\Omega = 4m^2\kappa_c^2 a^4$, indicating that it is only the S-wave (or $\ell = 0$) partial wave that participates. Physically this dominance of the S wave occurs for short-range interactions because a state with angular momentum ℓ behaves like r^ℓ near $r = 0$, so it is only the $\ell = 0$ mode that has a nonzero wave-function as $r \to 0$ and so can 'see' the short-range force. We expect from this that particles that are attracted through a short-range interaction — like nucleons in a nucleus — will like to pair up into $\ell = 0$ combinations if left to themselves.

3.3.5 Domain of validity of the Born approximation

We can (and should) ask when it is a good approximation to keep only the first terms in the Born series. For this we ask $|\psi_1|$ to be much smaller than $|\psi_0|$ and so on for higher corrections. Evaluating ψ_1 at $r = 0$ (where it should be biggest) and considering a short-range potential for which the integration range is only over $|\mathbf{x}| \sim a$, we expect $\psi_1 \sim ma^2 U\psi_0$, and so (putting back the $\hbar$ s on dimensional grounds) expect the approximation to work if $|\langle U\rangle| \ll \hbar^2/ma^2$, with the expectation taken in the initial state ψ_0. This says that the energy cost (imposed by the uncertainty principle) to be localized in the area of size a should be larger than the energy available in the potential there. (This seems reasonable given our experience with the square well, which shows that in this regime we do not expect bound states to exist — as might be expected in a perturbative regime.)

At high energies we get a weaker condition because we can profit from the explicit factor of $1/q \sim 1/k$ appearing in formulae like (3.125), that works to make it small. In this case the estimate $|\langle U\rangle| \ll 1/ma^2$ weakens to $|\langle U\rangle| \ll k/ma \sim v/a$. This says the frequency shift wrought by U should be much smaller than the inverse of the time taken to move across distances of order the range of the force.

[23] The agreement is eerie because we get the exact Coulomb result using only the leading Born approximation.

Exercise 3.10: The Yukawa potential, $U(r) = \kappa e^{-r/a}/r$, is a reasonable approximation to the force between nucleons (in which case $\kappa \simeq \mathcal{O}(1)$ and $a \simeq 1$ fm). What is the maximum nucleon kinetic energy that is consistent with the low-energy limit $ka \ll 1$? Should the Born approximation to work well for the force between nucleons at these energies? Are there any higher energies for which the Born approximation works while still trusting a non-relativistic treatment (take the nucleon mass to be 940 MeV)? Yukawa potentials are also a reasonable description of charged-particle scattering from a point charge that is screened by mobile opposite-sign charges (in which case $\kappa \simeq \alpha \simeq 1/137$ and a depends on the distances over which the mobile charges can move). Should we expect the Born approximation to work well for low-energy electron scattering from charges screened on atomic scales (for which $a \sim 0.1$ nm)? Should we expect it to work for low-energy muon scattering from a screened charge with the same value of a?

3.3.6 Scattering from charge distributions

Another useful application of the Born approximation is to the scattering from a continuous charge distribution, $\rho(\mathbf{x})$, rather than a point charge. In this case electrostatics tells us that the interaction potential with an incident point particle with charge Q becomes

$$U(\mathbf{r}) = \int \mathrm{d}^3\mathbf{x}\, \frac{\rho(\mathbf{x})}{|\mathbf{r}-\mathbf{x}|}\,. \tag{3.129}$$

What is relevant for scattering is the Fourier transform of this potential, which is

$$\tilde{U}(\mathbf{q}) = \int \mathrm{d}^3\mathbf{r}\, U(\mathbf{r})\, e^{-i\mathbf{q}\cdot\mathbf{r}} = \frac{4\pi}{q^2}\, \tilde{\rho}(\mathbf{q})\,, \tag{3.130}$$

where

$$\rho(\mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{l}}{(2\pi)^3}\, \tilde{\rho}(\mathbf{l})\, e^{-i\mathbf{l}\cdot\mathbf{x}}\,, \tag{3.131}$$

is the Fourier transform of the charge distribution. The differential cross section therefore becomes

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_C |F(\mathbf{q})|^2\,, \tag{3.132}$$

where $(\mathrm{d}\sigma/\mathrm{d}\Omega)_C$ denotes the Rutherford (point-particle) cross section from a point-particle carrying the same total charge,

$$Q_{\rm tot} = \int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x}) = \tilde{\rho}(\mathbf{q}=0)\,, \tag{3.133}$$

carried by the distribution.

The *form factor*, $F(\mathbf{q})$, is given by

$$F(\mathbf{q}) := \frac{\tilde{\rho}(\mathbf{q})}{Q_{\rm tot}}\ , \tag{3.134}$$

and clearly satisfies $F(0) = 1$. Because of this the extreme small-angle scattering (for which $q \to 0$) is not changed relative to the Rutherford result, which is reasonable since this occurs at such great impact parameters that it is sensitive only to the overall charge of the scatterer.

What $F(\mathbf{q})$ can change significantly, however, is the likelihood of large-angle scattering, for which q is not small. This allows us to quantify what would be expected for α scattering for a 'plum pudding' atom as Thompson envisioned it before Rutherford's experiment indicated the existence of a nucleus. Smoothly smearing the nuclear charge over the volume of an atom of radius $a \sim 1$ Angstrom (or 10^{-10} m) means that ρ is of order Ze/a^3, while $Q_{\rm tot} = Ze$ and $F(\mathbf{q})$ falls to zero quickly for $q \gtrsim 1/a$. Then the largest-angle scattering occurs when $q = 2k\sin(\theta/2) \sim 1/a$, so taking $k \sim 1/\lambda$ where $\lambda \sim 1$ fm (or 10^{-15} m) is a nuclear radius implies $\theta/2 \sim \sin(\theta/2) \lesssim \lambda/a \sim 10^{-5}$. No wonder Rutherford found his large-angle scattering result so surprising!

3.3.7 Multipole moments

Often experiments cannot probe the details of a charge distribution, such as if scattering happens not to probe sufficiently close to the charge distribution. In such cases it can happen that experiments are only sensitive to the lowest multipole moments of the distribution.

Expanding

$$\frac{1}{|\mathbf{r}-\mathbf{x}|} = \frac{1}{r}\left[1 - \frac{2\mathbf{r}\cdot\mathbf{x}}{r^2} + \frac{\mathbf{x}^2}{r^2}\right]^{-1/2} \simeq \frac{1}{r}\left[1 + \frac{\mathbf{r}\cdot\mathbf{x}}{r^2} - \frac{\mathbf{x}^2}{2r^2} + \frac{3}{2}\left(\frac{\mathbf{x}\cdot\mathbf{r}}{r^2}\right)^2 + \cdots\right], \qquad (3.135)$$

for $r \gg |\mathbf{x}|$, the long-distance form of the potential becomes

$$\begin{aligned} U(\mathbf{r}) &= \int \mathrm{d}^3\mathbf{x}\, \frac{\rho(\mathbf{x})}{|\mathbf{r}-\mathbf{x}|} \simeq \frac{1}{r}\int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x})\left\{1 + \frac{\mathbf{r}\cdot\mathbf{x}}{r^2} + \frac{1}{2r^2}\left[3\left(\frac{\mathbf{x}\cdot\mathbf{r}}{r}\right)^2 - \mathbf{x}^2\right] + \cdots\right\} \\ &= \left[\frac{Q_{\rm tot}}{r} + \frac{\mathbf{D}\cdot\hat{\mathbf{r}}}{r^2} + \frac{\hat{\mathbf{r}}\cdot\mathbf{Q}\cdot\hat{\mathbf{r}}}{2r^3} + \cdots\right], \end{aligned} \qquad (3.136)$$

where $\hat{\mathbf{r}} = \mathbf{r}/r$ is the usual radially pointing unit vector. This expression defines the total charge, $Q_{\rm tot}$, the *dipole moment* vector, $\mathbf{D}$, and *quadrupole moment* tensor, $\mathcal{Q}_{ij}$, as higher and higher moments of the charge distribution,

$$Q_{\rm tot} = \int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x}), \quad \mathbf{D} = \int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x})\, \mathbf{x}, \quad \mathcal{Q}_{ij} = \int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x})\left[3\, x_i\, x_j - \mathbf{x}^2\, \delta_{ij}\right], \qquad (3.137)$$

and so on. Only the first few are measurable if the size of the distribution is small enough that only a few powers of the ratio $|\mathbf{x}|/r$ are detectable. Notice that higher multipoles vanish, $D_i = \mathcal{Q}_{ij} = 0$, if ρ is a spherically symmetric distribution. A similar multipole expansion can also be defined for the magneto-static properties of a source, in analogy to the above discussion for electrostatics.

The multipole expansion is related to the small-$\mathbf{q}$ expansion of the form factor, as may be seen by Taylor expanding the Fourier transform of the charge distribution,

$$\begin{aligned}\tilde{\rho}(\mathbf{q}) &= \int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x})\, e^{i\mathbf{q}\cdot\mathbf{x}} = \int \mathrm{d}^3\mathbf{x}\, \rho(\mathbf{x}) \left[1 + i\mathbf{q}\cdot\mathbf{x} - \frac{1}{2}\,(\mathbf{q}\cdot\mathbf{x})^2 + \cdots\right] \\ &= Q_{\rm tot} + i\mathbf{D}\cdot\mathbf{q} - \frac{q^2}{6}\left[\hat{\mathbf{q}}\cdot\mathbf{Q}\cdot\hat{\mathbf{q}} + \langle\mathbf{x}^2\rangle\right] + \cdots\,, \end{aligned} \tag{3.138}$$

where $q = |\mathbf{q}|$ and $\langle\mathbf{x}^2\rangle = \int \mathrm{d}^3\mathbf{x}\ \rho(\mathbf{x})\,\mathbf{x}^2$. As we shall see, the fact that nuclei (and nucleons) have nontrivial multipole moments is part of the evidence that they are built from conglomerations of smaller charges.

4 Nucleon substructure

Historically, the first application of the tools just developed would be to winkle out the properties of nuclei and how this depends on the interactions between neutrons and protons. This turns out to be complicated, partly because the protons and neutrons themselves have substructure and so are complicated objects. Interactions amongst nucleons are complicated, much as are electromagnetic interactions among molecules, despite the simplicity of their main underlying root cause (for molecules, the Coulomb interaction).

This section therefore pauses the historical development and first develops the evidence for the compositeness of protons and neutrons, returning to nuclei in the next section.

What is an 'elementary' particle?

Before diving into the particulars about the substructure of protons, neutrons and nuclei it is first worth understanding what it means for a particle *not* to have substructure.

Colloquially, what is meant is intuitive: a particle is elementary when there is no evidence for it being built from constituents (as we shall see is the case at present for the electron). In practice what this means is that (so far as we can tell) the state of the particle is completely specified using only the values it holds for the small list of conserved quantum numbers: energy, momentum, angular momentum (or spin), electric charge and (perhaps) baryon number and the lepton numbers (if these really turn out to be conserved at a fundamental level). All other properties can be related to these basic ones, such as particle position which arises from taking linear superpositions of momentum eigenstates, as in $|\mathbf{x}\rangle \propto \int \mathrm{d}^3\mathbf{p}\, e^{i\mathbf{p}\cdot\mathbf{x}}\,|\mathbf{p}\rangle$, and so on.

Of course, what passes for elementary by this definition is usually a function of time, since as we learn more about a particle it can happen that evidence begins to emerge that more than these quantities are needed to specify its state completely. In the event, this is what happened with the atom, nucleus, proton and neutron, all of which initially were thought to be elementary until this was abandoned in the face of mounting evidence to the contrary.

4.1 Electrons, nucleons and quarks

The first step towards understanding nuclei is to understand the nature of the nuclear constituents: protons and neutrons (or *nucleons*, as they are both called collectively). Protons and neutrons, together with electrons, are the workhorses of atomic and nuclear structure, and their basic properties are summarized in Table 3. This reveals all three to be fermions and to have spin $\frac{1}{2}$ (not unrelated facts, as it turns out, because relativity and quantum mechanics together imply the *spin-statistics theorem* which forces spin-half particles to be fermions). The proton and electron have precisely opposite charge and all three have magnetic moments, as may be measured by observing their spins precess in an applied uniform magnetic field or by observing their motion in a spatially varying magnetic field.

Table 3. Properties of the electron, proton and neutron

Particle	statistics	rest mass	spin	charge	magnetic moment[a,b,c]	quark content
e^-	fermion	0.511 MeV	$\frac{1}{2}$	$-e$	$-1.00115965218073(28)\ \mu_B$	elementary
p	fermion	938 MeV	$\frac{1}{2}$	$+e$	$2.792847356(23)\ \mu_N$	uud
n	fermion	940 MeV	$\frac{1}{2}$	0	$-1.91304272(45)\ \mu_N$	udd

[a] The magnetic moment is proportional to spin, so what is given here is its value for a state with $S_z = +\frac{1}{2}$.

[b] The Bohr magneton is $\mu_B = e/2m_e$ and so $\mu_B = 5.7883818012(26) \times 10^{-5}$ eV/Tesla $= 9.27400968(20) \times 10^{-24}$ Joule/Tesla. In microscopic units $\mu_B = 193.0806\ e$-fm.

[c] The nuclear magneton is $\mu_N = e/2m_p$ and so $\mu_N = 3.1524512550(15) \times 10^{-8}$ eV/Tesla $= 5.05078353(11) \times 10^{-27}$ Joule/Tesla. Equivalently $\mu_N = 0.105155\ e$-fm.

4.1.1 Magnetic moments and quarks

The value of a particle's magnetic moment provides a clue about whether it has substructure. Although it is natural for a charged particle with spin to have a magnetic moment, the precise value of this moment turns out to be determined by its charge and mass, so comparison with this expectation provides a measure of whether or not the particle can be elementary.

To get an idea of the relationship between a particle's magnetic moment and spin consider a classical rigid body of total mass m and electric charge q that spins about an axis passing through its c.o.m. with angular velocity ω. Any infinitesimal volume element, $\mathrm{d}^3\mathbf{x}$, of the body carries a small part, $\mathrm{d}m = \rho(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$, of its mass and a small element $\mathrm{d}q = \sigma(\mathbf{x})\,\mathrm{d}^3\mathbf{x}$, of its charge. It is the motion of this element of charge that is responsible for the particle's magnetic moment, since its motion in a circle of radius r makes it into a small loop of area πr^2 about which an element $\mathrm{d}q$ of charge rotates once per period $\tau = 2\pi r/v$. This corresponds to an electrical current, $\mathrm{d}I = \mathrm{d}q/\tau = v\,\mathrm{d}q/2\pi r$.

Since the magnitude of the magnetic moment of a small current loop is the product of the current times the loop's area, we know that rotation of the volume element, $\mathrm{d}^3\mathbf{x}$, generates a magnetic moment of magnitude

$$\mathrm{d}\mu = (\pi r^2)\,\mathrm{d}I = (\pi r^2)(v\,\mathrm{d}q/2\pi r) = \frac{1}{2}\,rv\,\sigma(\mathbf{x})\,\mathrm{d}^3\mathbf{x}\,. \tag{4.1}$$

By contrast, the contribution of this same volume element to the magnitude of the body's angular momentum is

$$\mathrm{d}L = rv\,\mathrm{d}m = rv\,\rho(\mathbf{x})\,\mathrm{d}^3\mathbf{x}\,, \tag{4.2}$$

and so

$$\mathrm{d}\mu = \frac{1}{2}\left(\frac{\mathrm{d}q}{\mathrm{d}m}\right)\mathrm{d}L = \frac{q}{2m}\,\mathrm{d}L\,, \tag{4.3}$$

where the last equality assumes $\mathrm{d}q/\mathrm{d}m = \sigma(\mathbf{x})/\rho(\mathbf{x}) = q/m$ is $\mathbf{x}$-independent (*i.e.* that the charge and mass distribution are proportional to one another). Of course both $\boldsymbol{\mu}$ and $\mathbf{L}$ are vectors and particles are not classical rigid bodies. So the relation between magnetic moment and angular momentum is instead usually written[24]

$$\boldsymbol{\mu} = \frac{gq}{2m}\,\mathbf{s} = \gamma\,\mathbf{s}\,, \tag{4.4}$$

where $\mathbf{s}$ is the particle spin, the parameter g is called the *Landé g-factor* and $\gamma = gq/2m$ is its gyromagnetic ratio.

It turns out in the absence of interactions a spin-$\frac{1}{2}$ particle with no substructure should have magnetic moment component, $\mu_z = \pm q/2m$ where q is its charge and m is its mass, and so (because $s_z = \pm\frac{1}{2}$) for non-interacting spin-$\frac{1}{2}$ particles we expect $g = 2$ (or very close to this once interactions are included). (This value for the spin and magnetic moment is predicted by the *Dirac equation*, which we now understand describes fundamental spin-half particles once relativity is combined with quantum mechanics into quantum field theory.)

For the electron we have $s_z = \pm\frac{1}{2}$ and $q = -e$ and so $\mu_z = \mp g\mu_B/2 \simeq \mp\mu_B$ where $\mu_B = e/2m_e$ is called the *Bohr magneton* and we use $g \simeq 2$. It turns out that μ is very well-measured for electrons and so we know g is not precisely 2. The small deviations of g from 2 for the electron are well-understood as being due to small corrections from Quantum Electrodynamics (QED),[25] with $g - 2$ calculable (and calculated) perturbatively in a series

[24] This is the historical convention for g. More recently the definition of g has been defined using $\mu = (ge/2m)L$, using the proton charge even for the electron. In this convention g is negative for the electron rather than positive, as it is here.

[25] The corrections arise because the magnetic moment is defined by the energy shift of different spin states when a magnetic field is applied. But a magnetic field is itself really a quantum operator, $\hat{\mathbf{B}}$, which in QED does not commute with energy and so fluctuates in an energy eigenstate (in the same way that position or momentum fluctuate in the ground state of a harmonic oscillator). The extra 'anomalous' contribution to the magnetic moment is due to the electron's interacting with these fluctuations in the applied magnetic field.

in $\alpha/\pi \simeq 0.002$, where $\alpha = e^2/4\pi\hbar c$ is the fine-structure constant. The value for g obtained in this way is

$$g = 2\left[1 + \sum_{n=1}^{\infty} c_i \left(\frac{\alpha}{\pi}\right)^n\right] ,, \tag{4.5}$$

where $c_1 = \frac{1}{2}$ and many higher-order terms have also been computed (and are required when comparing to measurements due to the fantastic accuracy of the experiments). Writing $g = 2(1 + a)$ one finds for electrons:

$$\begin{aligned} a_e^{\rm exp} &= 11596521807.3(2.8) \times 10^{-13} \\ a_e^{\rm th} &= 11596521817.8(0.6)(0.4)(0.2)(7.6) \times 10^{-13} . \end{aligned} \tag{4.6}$$

The numbers in brackets represent the best present assessment of the error. For the theoretical calculation the first error arises from the uncertainty in our knowledge of the coefficient c_4 in the series expansion; the second error comes from the uncertainty in c_5; the third error is an uncertainty coming from difficulty with calculating with the strong interactions (which play a role at higher orders in α) and the fourth error comes from the uncertainty in the value of α itself. Clearly measurements agree with predictions and this agreement is one of the most precise agreements between theory and experiment known to physics. Such spectacular agreement is part of the evidence that electrons really do behave as if they were elementary.

The proton and neutron magnetic moments, on the other hand, provide evidence that (unlike electrons) these are *not* elementary and so likely do have substructure. We now know them each to be built from three constituents, called quarks. Two types of quarks, called 'up' and 'down' quarks (or u and d) appear in protons and neutrons, with the proton being built from the combination uud and the neutron from udd. The quarks also are fermions and carry spin $\frac{1}{2}$, and have electric charges $q_u = 2e/3$ and $q_d = -e/3$ (so that $q_p = 2q_u + q_d = e$ and $q_n = q_u + 2q_d = 0$). The quarks are themselves believed to be elementary, and their properties describe well the features of nucleons (and the rest of the zoo of strongly interacting particles — collectively called *hadrons* — besides nucleons).

Table 4. Some properties of u and d quarks.

Particle	statistics	'current' mass (m)	'constituent' mass (M)	spin	charge
u	fermion	2.3 ± 0.5 MeV	$\simeq 336$ MeV	$\frac{1}{2}$	$\frac{2}{3}\,e$
d	fermion	4.8 ± 0.5 MeV	$\simeq 340$ MeV	$\frac{1}{2}$	$-\frac{1}{3}\,e$

Some of the properties inferred for u and d quarks are given in Table 4, including two different notions of mass for these particles. (There can be more than one notion because quark masses are always inferred indirectly, since no isolated quark has ever been seen outside

a hadron.) What is listed as the 'current' mass, m, is most similar to what we normally mean by a rest mass: free quarks would satisfy $E^2 = \mathbf{p}^2 + m^2$, for each quark species, in the absence of interactions. By contrast, 'constituent' masses only arise because inside hadrons quarks are confined to live within a very small distance ($\ell \sim 1$ fm, set by the strong interactions that bind them there) whose value determines the physical size of particles like protons and neutrons. Because of this their momenta are bounded below by the uncertainty principle: $|\mathbf{p}| \geq |\mathbf{p}_{\rm min}| \simeq \pi/\ell$, and because $1/\ell \gg m$ we have $E \geq M := \sqrt{\mathbf{p}_{\rm min}^2 + m^2} \simeq |\mathbf{p}_{\rm min}| \simeq \frac{\pi}{\ell}$. For this reason it is M that often plays the role of mass, in the sense of being the smallest energy available to the quark as a function of its available momentum.

What binds quarks together so tightly that they can form nucleons? And why do nucleons contain three of them rather than just two? We shall see that this is a consequence of the underlying strong interactions which the quarks experience. Although more detail is given later, each quark turns out to come in three 'colours' (say, red, green and blue), so

$$u = \begin{pmatrix} u_{\rm red} \\ u_{\rm green} \\ u_{\rm blue} \end{pmatrix} , \tag{4.7}$$

and ditto for d. We are familiar with how electromagnetic interactions try to build bound states (like atoms) that are electrically neutral. They do so because if they do not, electrical forces remain in play that continue to attract things together. It is only once they are neutral that the electrical forces are neutralized (hence the name) and so no longer cause lower-energy configurations to be sought. In the same way, strong interactions turn out to try to build bound states that are *colour-neutral* in the sense that they are invariant under 3×3 unitary rotations of the quark colours

$$u \to U u \qquad \text{or} \qquad u_i = U_i{}^j \, u_j \tag{4.8}$$

where U is an arbitrary unitary matrix, called a colour 'rotation'. The second way of writing this makes explicit the three colour components, u_i, $i = 1, 2, 3$, of the quark, and $U_i{}^j$ is one of the elements of the matrix U, with i labelling the row and j the column of the entry. Also used is the Einstein summation convention which states that any repeated index must be summed over all of its allowed values. For instance, $U_i{}^j \, u_j$ denotes $\sum_{j=r,b,g} U_i{}^j \, u_j$ as required for matrix multiplication.

It turns out there are two ways of combining quarks into invariant, colour-neutral, combinations. One is to take the completely antisymmetric combination of the three quark colours. The other is to combine a quark and an antiquark. Each of these combinations corresponds to a known type of hadron: the 3-quark combinations are the *baryons* (which include the proton and neutron) and the quark-antiquark combinations are the *mesons* (which include $\pi^+ = u\overline{d}$, for example). Schematically these are written

$$B = \epsilon^{ijk} q_i q_j q_k \qquad \text{and} \qquad M = \overline{q}^i q_i \, , \tag{4.9}$$

Table 5. Ground-state mesons built from u and d quarks

Particle	spin	charge	isospin	mass	decay width	quark content
π^+	0	$+e$	1	140 MeV	$(2.6\times10^{-8}\text{ s})^{-1}$	$u\overline{d}$
π^-	0	$-e$	1	140 MeV	$(2.6\times10^{-8}\text{ s})^{-1}$	$\overline{u}d$
π^0	0	0	1	135 MeV	$(8.5\times10^{-17}\text{ s})^{-1}$	$(\overline{u}u\,,\overline{d}d)$
η	0	0	0	548 MeV	1.3 keV	$(\overline{u}u\,,\overline{d}d)^a$
ρ^+	1	$+e$	1	770 MeV	149 MeV	$u\overline{d}$
ρ^-	1	$-e$	1	770 MeV	149 MeV	$\overline{u}d$
ρ^0	1	0	1	770 MeV	149 MeV	$(\overline{u}u\,,\overline{d}d)$
ω	1	0	0	782 MeV	8.5 MeV	$(\overline{u}u\,,\overline{d}d)^a$

[a] Can also involve significant admixtures of $\overline{s}s$ (though less so for ω than η).

where the index i runs over the three values red, green and blue, and ϵ^{ijk} is the completely antisymmetric tensor that vanishes if any of its indices are equal, and otherwise takes values $\epsilon^{ijk} = +1$ (or -1) according to whether ijk is an even (or odd) permutation of $123 = \text{red}, \text{green}, \text{blue}$. Also q here denotes any quark species (*e.g.* either u or d for the present purposes).

These rules capture precisely the known hadrons. The situation is illustrated by Tables (5) and (6), which specialize to combinations involving only u and d quarks (in reality there are also four other known quark flavours: s, c, b and t). There should be four possible ways to combine quark flavours in a $\overline{q}q$ combination, and four possible ways to combine their spins.[26] As shown in Table (5) all such combinations are indeed observed (keeping in mind that each spin-1 particle has 3 spin-states). (The η meson also involves admixtures of other quarks, such as $\overline{s}s$ *etc.*, though this does not change the counting argument being made here.)

The baryons are captured equally well. For these Fermi statistics and the requirement of colour neutrality (which makes the colour part of the state already completely antisymmetric) say that a baryon must be completely symmetric under the simultaneous interchange of the spin and flavour of any pair of the 3 quarks. This can be done by symmetrizing separately for spin and for flavour — leading to the spin-$\frac{3}{2}$ combination corresponding to the 'Δ resonances' — or by combining states of mixed spin and flavour symmetry that are only symmetric once both are interchanged, corresponding to the spin-$\frac{1}{2}$ and isospin-$\frac{1}{2}$ nucleon.

Returning to magnetic moments, for some purposes it is useful to think of quarks with constituent masses as being non-relativistic within a proton and this gives a simple understanding of nucleon (and more generally, hadron) magnetic moments. In particular, since (as is often the case) the ground state dominantly has no orbital angular momentum, the nucleon

[26] We assume ground-state configurations to have zero orbital angular momentum, as is the case.

Table 6. Ground-state baryons built from u and d quarks

Particle	spin	charge	isospin	mass	decay width	quark content
p	1/2	$+e$	1/2	938 MeV	0	uud
n	1/2	0	1/2	940 MeV	$(880\text{ s})^{-1}$	udd
Λ^0	1/2	0	0	1116 MeV	$(0.26\text{ ns})^{-1}$	uds
Σ^-	1/2	$-e$	1	1197 MeV	$(0.1\text{ ns})^{-1}$	dds
Σ^0	1/2	0	1	1193 MeV	$(7\times10^{-20}\text{ s})^{-1}$	uds
Σ^+	1/2	$+e$	1	1189 MeV	$(0.08\text{ ns})^{-1}$	uus
Δ^-	3/2	$-e$	3/2	1232 MeV	117 MeV	ddd
Δ^0	3/2	0	3/2	1232 MeV	117 MeV	udd
Δ^+	3/2	$+e$	3/2	1232 MeV	117 MeV	uud
Δ^{++}	3/2	$+2e$	3/2	1232 MeV	117 MeV	uuu
Ξ^0	1/2	0	1/2	1315 MeV	$(0.3\text{ ns})^{-1}$	dss
Ξ^-	1/2	$-e$	1/2	1322 MeV	$(0.16\text{ ns})^{-1}$	uss
Ω^-	3/2	$-e$	0	1672 MeV	$(0.08\text{ ns})^{-1}$	sss

magnetic moment becomes the sum of those of the underlying quarks, so a straightforward calculation reveals nucleon moments are given by

$$\mu_p = \frac{4}{3}\,\mu_u - \frac{1}{3}\,\mu_d \qquad \text{and} \qquad \mu_n = \frac{4}{3}\,\mu_d - \frac{1}{3}\,\mu_u\,, \tag{4.10}$$

and these turn out to agree very well with the measured values when the quark moments are taken to have the fundamental form: $\mu_a = q_a/2M_a$. In particular, because the constituent masses of the u and d are very similar we expect $\mu_u/\mu_d \simeq q_u/q_d = -2$, and so

$$\frac{\mu_n}{\mu_p} = \frac{4-(\mu_u/\mu_d)}{4(\mu_u/\mu_d)-1} = -\frac{2}{3} \simeq -0.667 \tag{4.11}$$

which compares well with the experimental value $-1.913/2.793 \simeq -0.685$ given the accuracy of the relation $M_u \simeq M_d$. Given this, the absolute value of μ_p or μ_n essentially determines the average value of the constituent masses, M, as listed in the table.

But the power of the idea of compositeness is that it is predictive rather than just being descriptive. In particular the same calculation should not be limited only to protons and neutrons, but should also give the magnetic moment of all baryons in terms of quark moments. This leads to the (successful) predictions given in Table 7.

4.2 Elastic scattering

Much of what we know about proton and neutron substructure comes from scattering experiments, particularly those where the nucleon is probed using particles that seem to have

Table 7. Baryon magnetic moments predicted by the quark model

Particle	spin	charge	quark content	$\mu_{\rm th}$
p	1/2	$+e$	uud	$\frac{1}{3}(4\mu_u - \mu_d)$
n	1/2	0	udd	$\frac{1}{3}(4\mu_d - \mu_u)$
Λ^0	1/2	0	uds	μ_s
Σ^-	1/2	$-e$	dds	$\frac{1}{3}(4\mu_d - \mu_s)$
Σ^0	1/2	0	uds	$\frac{1}{3}(2\mu_u + 2\mu_d - \mu_s)$
Σ^+	1/2	$+e$	uus	$\frac{1}{3}(4\mu_u - \mu_s)$
Ξ^0	1/2	0	uss	$\frac{1}{3}(4\mu_s - \mu_u)$
Ξ^-	1/2	$-e$	dss	$\frac{1}{3}(4\mu_s - \mu_d)$
Ω^-	3/2	$-e$	sss	$3\mu_s$

no substructure themselves, like electrons. For this reason it is useful first to summarize how things look when electrons scatter from another point-like particle, since this provides the point-particle benchmark against which to compare. For technical reasons we compare electron-proton elastic scattering to electron-muon scattering, rather than to electron-electron scattering. Electron-muon scattering is simpler than electron-electron scattering for these purposes for two reasons: (*i*) some of the weak interactions contribute differently for ee vs $e\mu$ (or ep) scattering; and (*ii*) even the purely electromagnetic part of ee scattering is complicated by the identical nature of the initial and final particles.

> **Exercise 4.1:** The magnetic moment of the Ω^- baryon is measured to be $\mu_\Omega = -2.02 \pm 0.05\ \mu_N$. Use this and the values given in Table 3 together with the predictions of Table 7 to infer the values of the magnetic moments for the u, d and s quarks. Assuming each of these quarks, $\{q_i\} = \{u, d, s\}$, has a dipole moment of size $Q_i/(2M_i)$ where Q_i and M_i are the quark charge and *constituent* mass, what do these values for magnetic moments imply about the constituent mass for each of the u, d and s quarks? Compute the predictions for the numerical values of the rest of the hadronic magnetic moments given in Table 7.

4.2.1 Elastic $e\mu$ scattering

When electrons elastically collide with muons their collisions *are* well-described by point-particle scattering, and this is a large part of why we believe both the electron and the muon to be elementary. So far as their electromagnetic interactions are concerned, muons are pretty much identical to electrons except for the fact that they have different masses: $m = m(e) = 0.511$ MeV and $M = m(\mu) = 106$ MeV. The electromagnetic scattering of two point-like, unpolarized spin-$\frac{1}{2}$ particles of masses m and M has the invariant squared

amplitude

$$\mathcal{M}_{\rm em}(e\mu\to e\mu) = \frac{32\pi^2\alpha^2}{t^2}\Big[(m^2+M^2-u)^2+(s-m^2-M^2)^2+2t(m^2+M^2)\Big]$$
$$= \frac{32\pi^2\alpha^2}{(k-k')^4}\Big[(2k\cdot p')^2+(2k\cdot p)^2-2(m^2+M^2)(k-k')^2\Big]\,, \qquad (4.12)$$

where, as before, $\alpha = e^2/4\pi\hbar c$ is the electromagnetic fine-structure constant, and the Mandelstam variables are given by the usual expressions in terms of the energy-momentum 4-vectors:[27] $e(k)+\mu(p)\to e(k')+\mu(p')$.

Exercise 4.2: Working at energies much larger than the electron mass, m, (but not necessarily larger than $M \gg m$) neglect $m \simeq 0$ and use (4.12) to derive the differential cross section for $e\mu$ electromagnetic elastic scattering:

$$\frac{{\rm d}\sigma}{{\rm d}u{\rm d}t} = -\frac{4\pi\alpha^2}{\xi(s)f(s)}\left[\frac{(M^2-u)^2+(s-M^2)^2+2tM^2}{t^2}\right]\delta(s+t+u-2M^2)\,, \quad (4.13)$$

where (2.45) and (2.65) tell us

$$f = -4v_{\rm rel}(p\cdot k) = 4\sqrt{(p\cdot k)^2-m^2M^2} \simeq -4p\cdot k \qquad \text{and} \qquad \xi(s)\simeq s-M^2\,. \qquad (4.14)$$

In the lab frame take the initial muon to be at rest, the direction of the initial electron to be along the z axis and the plane of the scattering to be the $x-z$ plane. Show that an appropriate choice for the x, y and z axes (and 3-momentum conservation) implies the four 4-momenta in the reaction can be written

$$p^\mu = \begin{pmatrix} M\\0\\0\\0\end{pmatrix}\,,\; k^\mu = \begin{pmatrix}\omega\\0\\0\\\omega\end{pmatrix}\,,\; p'^\mu = \begin{pmatrix} E\\-\omega'\sin\theta\\0\\\omega-\omega'\cos\theta\end{pmatrix}\,,\; k'^\mu = \begin{pmatrix}\omega'\\\omega'\sin\theta\\0\\\omega'\cos\theta\end{pmatrix}\,, \qquad (4.15)$$

where E and ω' are determined by $E^2 = (\omega'\sin\theta)^2+(\omega-\omega'\cos\theta)^2+M^2 = \omega^2+(\omega')^2-2\omega\omega'\cos\theta+M^2$ and $\omega+M=\omega'+E$. Use these to prove the useful formulae

$$\omega' = \frac{\omega}{1+2(\omega/M)\sin^2(\theta/2)}$$
$$t = -4\,\omega\omega'\sin^2\frac{\theta}{2} \qquad \text{and} \qquad {\rm d}t = -2(\omega')^2\sin\theta{\rm d}\theta \qquad (4.16)$$
$$k\cdot p = -\omega M \qquad \text{and} \qquad k\cdot p' = k'\cdot p = -\omega' M\,,$$

[27] Here the notation $e(k)$ means the 4-momentum of e is denoted k^μ, and so on. Consequently energy-momentum conservation for the reaction indicated is the 4-vector condition: $k+p=k'+p'$.

and so that

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq \frac{\alpha^2}{8\omega^2}\left[1+\left(\frac{\omega'}{\omega}\right)^2 - 2\left(\frac{\omega'}{\omega}\right)\sin^2\frac{\theta}{2}\right]\csc^4\frac{\theta}{2} \qquad \text{(lab frame, } m_e = 0\text{)}\,. \tag{4.17}$$

Use (4.16) to show that this can be rewritten in an equivalent, often-used, form

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq \frac{\alpha^2}{4\omega^2}\frac{\cos^2(\theta/2)}{\sin^4(\theta/2)}\left(\frac{\omega'}{\omega}\right)\left[1+\left(\frac{2\omega\omega'}{M^2}\right)\sin^2\frac{\theta}{2}\tan^2\frac{\theta}{2}\right] \qquad \text{(lab frame, } m_e = 0\text{)}\,. \tag{4.18}$$

The limit where $\omega/M \to 0$ corresponds to the case where the target particle is so heavy it does not recoil and so the lab frame and the c.o.m. frame coincide. In this case $\omega' \to \omega$, so the electron energy does not change, and the cross section reduces to the result for a spinning electron scattering from a Coulomb potential, called the *Mott* scattering cross section:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}} = \left(\frac{\alpha^2}{4\omega^2}\right)\frac{\cos^2(\theta/2)}{\sin^4(\theta/2)}\,. \tag{4.19}$$

The factor of $\cos^2(\theta/2)$ arises due to the interaction between the moving electron magnetic moment and the Coulomb potential (which in the electron rest frame has a magnetic component). The prefactor is precisely the Rutherford result (with $U = \pm\alpha/r$) for the scattering of a spinless particle from a Coulomb potential.

4.2.2 Elastic *ep* scattering

What about *ep* scattering? We first look at elastic scattering, in which the incoming particle scatters collectively from the entire proton without transferring energy in the c.o.m. frame. In this case the outcomes of the experiments turn out not to be well-described by point-particle scattering and instead are better described if point-particle scattering is modified by form factors (along the lines described for a charge distribution in the previous section).

For the scattering of a spinless charged particle where the initial projectile energy is much less than the target mass (in the target rest frame), $\omega \ll M$, the scattering should be equivalent to Coulomb scattering from some sort of charge distribution. In this case we would write

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \simeq \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}} |F(\mathbf{q})|^2 \qquad (m_e \ll \omega \ll M, \text{ lab frame})\,, \tag{4.20}$$

where we've seen that the form factor, $F(\mathbf{q})$, is (up to normalization) the Fourier transform of the proton's internal charge distribution, and the goal is to extract its shape given scattering measurements, normalized so that $F(0) = 1$.

Unfortunately the electron spin prevents this last expression from being directly used for relativistic electron scattering. Instead there is a separate form factor for the electron's electric

and magnetic couplings, respectively called $G_E(q^2)$ and $G_M(q^2)$, where both are Lorentz-invariant functions of the electron's 4-momentum transfer: $q^\mu = k^\mu - (k')^\mu$. Because they are Lorentz-invariant the functions G_E and G_M can only depend on q^μ through its invariant length: $q^2 = \eta_{\mu\nu} q^\mu q^\nu$. But this is a variable we have seen before: it is one of the familiar Mandelstam invariants: $q^2 = -t$, which the above calculation of $e\mu$ scattering shows is given in the target rest-frame by $q^2 = -t = 4\,\omega\omega' \sin^2\frac{\theta}{2}$, where ω and ω' are the initial and final electron energies.

In the same way as for potential scattering, the values of G_E and G_M at $q^2 = 0$ are related to the proton's static electric charge and magnetic moment. In particular, $eG_E(0)$ is the proton's electric charge and so we know $G_E(0) = 1$. Similarly, the proton's magnetic moment turns out to be given by $\mu_p = \mu_N G_M(0)$, and so $G_M(0) - 1$ describes the deviation from the noninteracting point-particle result, $\mu_p = \mu_N$, where $\mu_N = e/2M = e/2m_p$ is called the *nuclear magneton*. The point-particle limit corresponds to the choice $G_E(q^2) = G_M(q^2) = 1$.

Once written in terms of G_E and G_M, the cross section for elastic ep scattering turns out to be given by the *Rosenbluth formula*,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}} \frac{\omega'}{\omega}\left[\frac{G_E^2 + \tau\, G_M^2}{1+\tau} + 2\tau\, G_M^2\, \tan^2\frac{\theta}{2}\right], \tag{4.21}$$

where M now denotes the proton mass and

$$\tau = \frac{q^2}{4M^2} = -\frac{t}{4M^2} = \frac{\omega\omega'}{M^2}\,\sin^2\frac{\theta}{2}\,. \tag{4.22}$$

Notice that this agrees with (4.18) when $G_E = G_M = 1$.

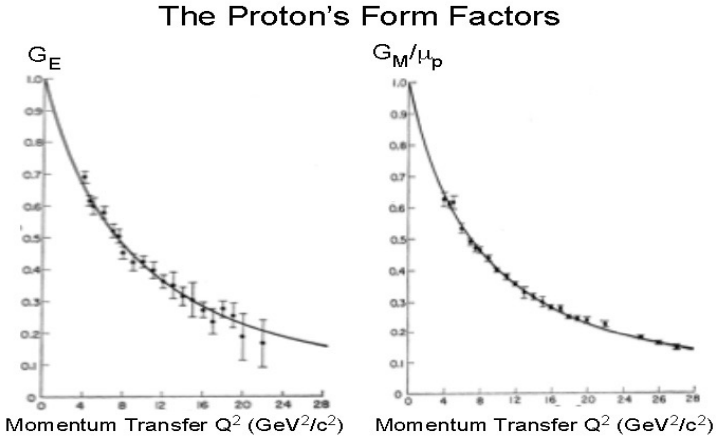


Figure 23. Measured values for the electric and magnetic proton form factors for elastic ep scattering. (Figure source: `http://www.mit.edu/~schmidta/olympus/guide.html`).

Figure 23 shows the form factors obtained by fits to elastic scattering experiments, which for ep scattering are performed by scattering electron beams from Hydrogen targets. (*en*

scattering is done using Deuterium targets, after subtracting out the *ep* contribution.) The results are clearly inconsistent (in both cases) with the point-proton limit $G_{_E} = G_{_M} = 1$. For protons the resulting functional form for small q^2 is not too far from the 'dipole' shape, wherein both form factors have the same dependence on q^2,

$$G_{_E}(q^2) \simeq \frac{1}{(1 + q^2/q_0^2)^2}\,, \tag{4.23}$$

where fits to experiments imply $q_0^2 \simeq 0.71$ (GeV/c)^2, and similarly for $\mu_{_N} G_{_M}/\mu_p$. As we have seen, this corresponds in position space to an exponential (or screened) charge distribution, $\sigma \propto e^{-r/a}/r$, with $a = 1/q_0 \simeq 0.23$ fm. Measurements of form-factors like these for small momentum transfer also determine the (electric and magnetic) mean effective radius of the nucleon, through expressions like

$$F(\mathbf{q}) = \int \mathrm{d}^3\mathbf{x}\; \rho(\mathbf{x})\, e^{-i\mathbf{q}\cdot\mathbf{x}} = 1 - \frac{\mathbf{q}^2}{6}\,\left\langle r^2 \right\rangle + \cdots\,. \tag{4.24}$$

Elastic scattering experiments such as these led to a picture of the proton as a mushy charge distribution, but did not directly point towards the nature of their substructure. It was the study of inelastic scattering, in which the collision is hard enough to disrupt the proton's internal structure, that provided much of the evidence that laid bare the role of quarks and gluons.

4.3 Inelastic *ep* scattering

Inelastic scattering occurs when some of the incoming electron energy is used to excite some internal degree of freedom of the target. For low energies this might start with the collisional excitation of the target to one of its excited energy levels, indicated by the appearance of a resonance in the cross section at the energy where $\omega - \omega' = E_{\rm res} - M$. At higher energies one might see internal constituents knocked out, such as occurs in electron-nuclear collisions when nucleons are kicked out of the initial nucleus. Still higher energies can see particle-anti-particle pair production and other interactions, possibly leading to very complicated many-particle final states.

In this section we examine a specific very informative process called *deep-inelastic scattering* in which electrons collide with nucleons at high enough energies to produce a wide variety of particles. Our interest, however, is not in the precise types of new particles that are produced, so we consider *inclusive* cross sections in which one only measures the energy and direction of the initial and final electron. The reaction is denoted $ep \to eX$ where X consists of any kind of hadronic final states caused by the disruption of the target nucleon, whose detailed properties are not measured so we sum over all possible final options for X.

Because we follow only the electron properties, the kinematic variables are very similar to what they were for elastic electron scattering. The main difference is that we no longer know

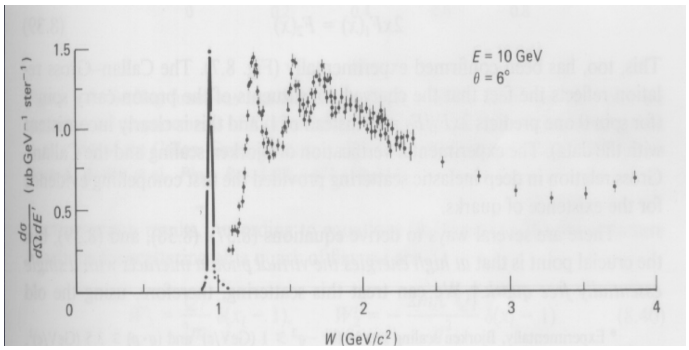


Figure 24. Sample *ep* double-differential scattering cross section, including the elastic peak (scaled down by a factor of 8.5 to be visible), several resonances and the beginnings of the continuum. (Figure source: `Annual Review of Nuclear and Particle Science, Volume 22 (1972) page 210.`)

the final rest mass of the total 4-momentum associated with X. That is, in $e(k) + p(p) \to e(k') + X(p')$ there is no longer a constraint that says $W^2 := -(p')^2 = (E')^2 - (\mathbf{p}')^2 = m^2$ where m is the mass of some specific particle. Instead we must regard W^2 as a kinematic variable whose final value in the cross section should be integrated over. Related to this is the fact that the final electron energy, ω', is no longer dictated by energy conservation purely as a function of ω and θ. So rather than following the dependence of the differential cross section, $\mathrm{d}\sigma/\mathrm{d}\theta$, as a function of two independent variables (θ and the initial energy ω), we instead imagine tracking the double-differential cross section, $\mathrm{d}\sigma/\mathrm{d}\omega'\mathrm{d}\theta$ as a function of the three independent variables θ, ω' and ω.

As usual it is useful to express the cross section in terms of manifestly relativistic variables, and for this we introduce the new energy-loss variable

$$\nu := -\frac{p \cdot q}{M}\,, \tag{4.25}$$

in addition to the familiar Mandelstam variable, $q^2 = (k - k')^2 = -t = 4\omega\omega' \sin^2 \frac{\theta}{2}$ (in which the final expression neglects the electron mass). The definition of ν reduces to $\nu = \omega - \omega'$ in the rest-frame of the target (whose mass we take, as before, to be M).

Similar to what happens for elastic scattering, the differential cross section for deep-inelastic scattering can also be written in terms of two form factors that parameterize our ignorance of the target nucleon's structure. The new feature in this case is that these form factors now can depend on *both* of the Lorentz-invariant variables q^2 and ν:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}q^2\mathrm{d}\nu} = \frac{4\pi\alpha^2}{Mq^4}\left(\frac{\omega'}{\omega}\right)\left\{W_2(q^2,\nu) + \left[2W_1(q^2,\nu) - W_2(q^2,\nu)\right]\sin^2\frac{\theta}{2}\right\}\,, \tag{4.26}$$

rather than just depending on q^2. In the special case of elastic scattering (for which X is the same as the target) ν is no longer independent of q^2, since then $M^2 = -(p')^2 = -(p+k-k')^2 =$

$M^2 - q^2 - 2p \cdot q$ and so $\nu = -p \cdot q/M = q^2/2M$, and in this limit W_1 and W_2 are related to the form factors $G_E(q^2)$ and $G_M(q^2)$.

In principle measurements of the cross section determine the $W_i(q^2, \nu)$, and thereby tell us about the substructure of the nucleon. To see how, imagine that at very high energies deep-inelastic scattering can be regarded as the incoherent hard scattering of the incident electron from one of the constituent quarks within the nucleon. If the quarks are themselves spin-$\frac{1}{2}$ point particles (as seems to be the case) then the electron-quark scatterings are themselves elastic, and so described by the $e\mu$ scattering cross section given earlier, (4.13):

$$\frac{\mathrm{d}\hat{\sigma}_a}{\mathrm{d}q^2} = -\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}t} = \frac{2\pi\alpha^2}{q^4}\left(\frac{\hat{e}_a^2}{e^2}\right)\left[\frac{(\hat{s}-q^2)^2 + \hat{s}^2}{\hat{s}^2}\right], \tag{4.27}$$

where $\hat{e}_a$ is the electric charge of quark type 'a' and we neglect both electron and quark masses at the energies of interest. We use 'hats' to denote 4-momenta and cross sections involving the quarks — sometimes also collectively called *partons*, together with the gluons inside the nucleon — to distinguish them from the corresponding quantities for the entire nucleon. So $\hat{\sigma}_a$ is the cross section for elastic electron scattering from quark type 'a' while $\hat{s} \simeq -2\hat{p} \cdot k$ is the Mandelstam variable computed using the initial quark 4-momentum, $\hat{p}^\mu$, rather than the 4-momentum of the entire target nucleon.

How is $\hat{s}$ related to s? For ultra-relativistic scattering ($\omega \gg M$) we can neglect both the nucleon mass and any zero-point energy of the quarks due to their being bound within the nucleon. We can therefore regard the quark 4-momentum as being parallel to the 4-momentum of the initial nucleon, $\hat{p}^\mu = x\, p^\mu$, with $0 \le x \le 1$ measuring the fraction of the initial nucleon's momentum carried by the quark in question. Consequently $\hat{s} \simeq -2\hat{p} \cdot k \simeq -2x\, p \cdot k \simeq xs$. But x is also related to ν because at the parton level the scattering is elastic, so

$$0 \simeq m_a^2 = -(\hat{p}')^2 = -(\hat{p} + k - k')^2 = -(x\, p + q)^2 \simeq -2x\, p \cdot q - q^2 = 2x\, \nu M - q^2\,, \tag{4.28}$$

and so

$$x = \frac{q^2}{2M\nu}\,. \tag{4.29}$$

The quantity x defined in this way is often called the *Bjorken scaling variable*, or *Bjorken x*.

To relate $\mathrm{d}\hat{\sigma}_a/\mathrm{d}q^2$ to the total electron-nucleon cross section we imagine there being a probability density, $\mathcal{P}_a(x)$, of finding quark type 'a' carrying a fraction x of the initial nucleon's momentum. For incoherent scattering the total interaction cross section therefore

becomes the sum over the cross section for scattering from each constituent quark, and so

$$\begin{aligned}
\frac{\mathrm{d}\sigma}{\mathrm{d}q^2\mathrm{d}\nu} &= \sum_a \int_0^1 \mathrm{d}x\, \mathcal{P}_a(x)\, \frac{\mathrm{d}\hat{\sigma}_a}{\mathrm{d}q^2}(x,q^2,s)\, \delta\left(\nu - \frac{q^2}{2Mx}\right) \\
&= \frac{2\pi\alpha^2}{q^4}\, \mathcal{P}(x)\, \left\{\left[1-\left(\frac{q^2}{xs}\right)\right]^2 + 1\right\} \frac{2Mx^2}{q^2}\bigg|_{x=q^2/2M\nu} \\
&= \frac{2\pi\alpha^2}{q^4}\, \mathcal{P}(q^2/2M\nu)\, \left[1+\left(\frac{\omega'}{\omega}\right)^2\right] \frac{2\omega\omega'\sin^2(\theta/2)}{M(\omega-\omega')^2} \\
&= \frac{4\pi\alpha^2}{Mq^4}\left(\frac{\omega'}{\omega}\right) \mathcal{P}(q^2/2M\nu)\, \left(1+\frac{2\omega\omega'}{\nu^2}\right) \sin^2\frac{\theta}{2} \\
&= \frac{4\pi\alpha^2}{Mq^4}\left(\frac{\omega'}{\omega}\right) \mathcal{P}(x)\, \left(\sin^2\frac{\theta}{2} + \frac{xM}{\nu}\right)\bigg|_{x=q^2/2M\nu},
\end{aligned} \tag{4.30}$$

where we repeatedly use $x = q^2/2M\nu$ and $q^2 = 4\omega\omega'\sin^2\frac{\theta}{2}$. Furthermore, we define

$$\begin{aligned}
\mathcal{P}(x) &:= \sum_{a=u,d,s,\bar{u},\bar{d},\bar{s}} \frac{\hat{e}_a^2}{e^2}\, \mathcal{P}_a(x) \\
&= \frac{4}{9}\Big[\mathcal{P}_u(x) + \mathcal{P}_{\bar{u}}(x)\Big] + \frac{1}{9}\Big[\mathcal{P}_d(x) + \mathcal{P}_{\bar{d}}(x) + \mathcal{P}_s(x) + \mathcal{P}_{\bar{s}}(x)\Big],
\end{aligned} \tag{4.31}$$

and recognize when performing the sum that the initial nucleon can contain a sea of $q\bar{q}$ pairs (or 'sea' quarks) in addition to the 'valence' uud or udd quarks and so include antiquarks in the sum on a. The quark sum is easy to do since the antiquark-electron scattering cross section is also given by (4.27) in the ultra-relativistic limit. It is a bit more complicated also to include the heaviest quarks, like b and t, since for these the quark masses need no longer be negligible, but the contributions of such heavy sea quarks to nucleon properties are usually negligible in practice.

> **Exercise 4.3:** Starting from eq. (4.27) and assuming that electron proton scattering is well-described by incoherent scattering from the proton's constituent quarks (whose masses can be neglected), derive the first line of eq. (4.30). Then use the result $\nu = \omega - \omega'$ in the proton rest frame, as well as relations (4.16), to derive the last line of eq. (4.30) with $\mathcal{P}(x)$ as given in (4.31).

When performing the sums we normalize the probabilities so that the integral over all x counts the number of valence quarks. For example, for protons (*i.e.* uud states) we have

$$\int_0^1 \mathrm{d}x\, \Big[\mathcal{P}_u(x) - \mathcal{P}_{\bar{u}}(x)\Big] = 2\,, \quad \int_0^1 \mathrm{d}x\, \Big[\mathcal{P}_d(x) - \mathcal{P}_{\bar{d}}(x)\Big] = 1\,, \quad \int_0^1 \mathrm{d}x\, \Big[\mathcal{P}_s(x) - \mathcal{P}_{\bar{s}}(x)\Big] = 0$$

and so on, while for neutrons (*i.e.* udd states) we instead have

$$\int_0^1 \mathrm{d}x\, \Big[\mathcal{P}_u(x) - \mathcal{P}_{\bar{u}}(x)\Big] = 1\,, \quad \int_0^1 \mathrm{d}x\, \Big[\mathcal{P}_d(x) - \mathcal{P}_{\bar{d}}(x)\Big] = 2\,, \quad \int_0^1 \mathrm{d}x\, \Big[\mathcal{P}_s(x) - \mathcal{P}_{\bar{s}}(x)\Big] = 0\,.$$

Notice that the sum in (4.31) could also have included a sum over *gluons* (the quanta for the strong-interaction force) but does not because these are electrically neutral and so do not take part in electromagnetic scattering. They *do* however carry some of the initial nucleon's momentum and because of this one typically finds that

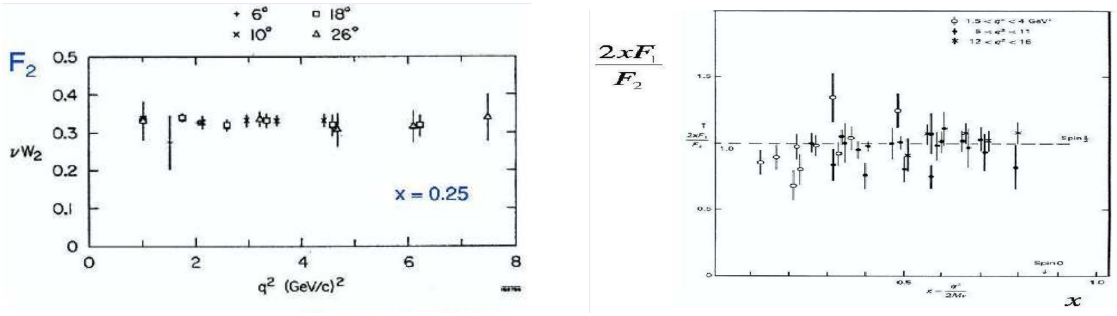
$$\sum_{a=q,\bar{q}} \int_0^1 \mathrm{d}x \; x\, \mathcal{P}_a(x) \simeq 0.5\,, \tag{4.32}$$

so that on average quarks carry only about half of a nucleon's initial momentum while the gluons carry the rest.

Because the form-factor expression, (4.26), is completely general (for electromagnetic scattering), it must include the parton-level calculation (4.30) as a special case. Comparing the two expressions allows the determination of W_1 and W_2, leading (in the limit $\nu = \omega - \omega' \gg M$) to the predictions $W_1(q^2,\nu) = F_1(x)$ and $\nu W_2(q^2,\nu)/M = F_2(x)$ with

$$2\,F_1(x) \simeq \mathcal{P}(x) + \mathcal{O}(M/\nu) \qquad \text{and} \qquad F_2(x) = x\,\mathcal{P}(x)\,. \tag{4.33}$$

These predictions for W_i agree well with experiments in the limit $\nu,\, \sqrt{q^2} \gg M$, and are in practice how the *quark distribution functions*, $\mathcal{P}_a(x)$, are determined. For example, an experimental test that F_2 depends only on x and not also on q^2 is shown in Figure 25. Also shown is a test of the prediction $2x\,F_1(x) = F_2(x)$ — called the *Callan-Gross relation* — that tests the spin-half nature of quarks since (for example) $F_1 = 0$ for spinless quarks. Both predictions are seen to be verified experimentally.



(a) q^2-independence of F_2 (b) Test of the Callan-Gross relation

Figure 25. Experimental tests of the parton picture of deep-inelastic electron-nucleon scattering. Left panel: demonstration that the structure function F_2 does not depend on q^2, which provides evidence that quarks themselves have no structure. Right panel: test of the Callan-Gross relationship that plots the measured ratio $2xF_1/F_2$ against Bjorken-x. This relation probes the spin-half nature of quarks. (Figure source: `D.H. Perkins, ''Introduction to High Energy Physics,'' Addison Wesley, 1987.`)

We see that the hypothesis that nucleons are made of point-like partons (quarks and gluons) agrees well with observations, and in many ways the experiments that showed this

are modern analogs of Rutherford's scattering experiment. We saw how ignorance of nucleon structure can be parameterized by structure functions, and elastic scattering from charge distributions gives these functions a strong dependence on q^2, whereas point particles would predict them to be q^2-independent. Although this strong q^2 dependence is indeed seen for low-energy scattering, in which the incident electron coherently scatters from the entire nucleon, it disappears again for very hard inelastic scattering. This reveals how very energetic electrons instead scatter dominantly from point-like quark constituents rather than from the proton as a whole, because the proton does not have time to respond to the delivered momentum transfer.

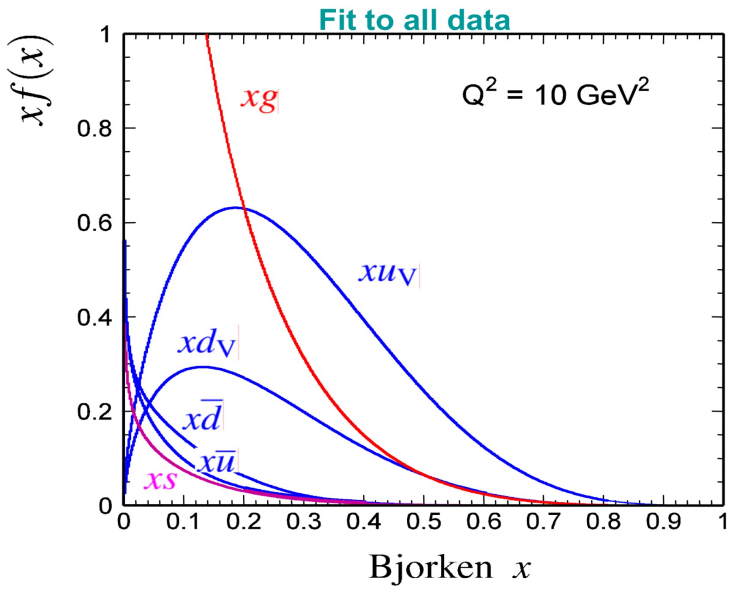


Figure 26. Plots of the parton distribution functions, $x\,\mathcal{P}_a(x)$, for several parton species as obtained by fits to multiple scattering processes. Valence quarks are seen to dominate for large x while sea quarks and especially gluons become more important for smaller x. (Figure source: `http://www.hep.phy.cam.ac.uk/ thomson/lectures/lectures.html`.)

What is important in all this is that once the functions, $\mathcal{P}_a(x)$, are determined from measurements (such as from *ep* scattering) they then can be used to predict *any* other kind of hard nucleon scattering by point objects. This is because the distribution of quarks within the nucleon is an intrinsic property of the nucleon, and does not care what particles are used to scatter against it in a particular experiment. So precisely the same functions also appear in neutrino-nucleon scattering, for example, though weighted by different coupling constants due to the different interactions involved. This is what makes these calculations predictive, given that we ultimately must obtain the functional form for $\mathcal{P}_a(x)$ from experiments. Plots of the values of $x\,\mathcal{P}_a(x)$ for different types of partons are shown in Figure 26, as obtained by

fitting the results of many collision processes.

Exercise 4.4: Suppose the form factors $F_i^{ep}(x)$ and $F_i^{en}(x)$ (with $i = 1, 2$) are measured for inclusive deep inelastic scattering for both $e^- + p \to e^- + X$ and $e^- + n \to e^- + X$ collisions. Assuming that electromagnetic interactions dominate and assuming that the antiquark distributions are equal within the proton and neutron – *i.e.* that $\mathcal{P}_{\bar{u}} \simeq \mathcal{P}_{\bar{d}} \simeq \mathcal{P}_{\bar{s}}$ is the same functions for both protons and neutrons – derive the Gottfried sum rule, which states:

$$\int_0^1 \frac{\mathrm{d}x}{x} \Big[F_2^{ep}(x) - F_2^{en}(x) \Big] = \frac{1}{3} \,.$$

5 Nuclear structure

We now return to much lower energies (several MeV) than in the higher energies $E \gtrsim 200$ MeV we've seen to be associated with the substructure of the proton and neutron in previous section. The goal in this section is to discuss the properties of how nucleons organize themselves into nuclei and see what this can tell us about the nature of the strong force that acts between nucleons.

The central principle of nuclear physics is the idea that atomic nuclei are bound systems built from mutually interacting protons and neutrons. Much of the evidence for this comes from experiments which collide protons and neutrons with nuclei or from radioactive decays. Numerous reactions, like $n + {}^3\mathrm{He} \to {}^4\mathrm{He} + \gamma$ or $n + {}^{14}\mathrm{C} \to {}^{14}\mathrm{N} + p$, show how individual nucleons can both be absorbed or ejected as one type of nucleus is transformed into another.

The picture that nuclei are built from protons and neutrons ultimately means that nuclear properties should be completely characterized by the number Z of protons and the number N of neutrons the nucleus contains, allowing (in principle) quantities like binding energies and the nature of any excited states to be deduced from these. It is common in practice to trade N for the total number of nucleons in the nucleus: $A := N + Z$. Because Z determines the nuclear electric charge it also determines how many electrons must bind to it to get a neutral atom, and therefore also determines the atom's chemical properties.

A reminder about notation: because the number Z determines chemical properties it is often specified only implicitly by identifying the relevant chemical element: *i.e.* a nucleus with $Z = 2$ protons and $N = 1$ neutrons has $A = 3$ and can be equivalently described as the nucleus with $(Z, A) = (2, 3)$ or as ${}^3\mathrm{He}$, since atoms for the element Helium all have $Z = 2$. Similarly ${}^{12}\mathrm{C}$ denotes a Carbon nucleus which has $(Z, A) = (6, 12)$. The superscript A is written to the left to allow room on the right to put in electric charge – as in ${}^3\mathrm{He}^+$ for the singly-charged Helium ion. Nuclei that share the same value of Z but different values of A are called nuclear *isotopes* for the element in question (*e.g.* ${}^{12}\mathrm{C}$ and ${}^{14}\mathrm{C}$ are isotopes of Carbon that respectively have $(Z, A) = (6, 12)$ and $(Z, A) = (6, 14)$).

As we shall see, the nuclear structure to which this leads is more complicated than is the electronic structure of an atom, for several reasons. The first reason is because the atom is dominated by the large, massive central nucleus to which the very light electrons respond through the long-ranged and well-understood Coulomb interaction. By contrast, all the constituents of a nucleus have similar mass and they are bound by what turns out to be a very strong but short-ranged force.

A second reason nuclei are more complicated than atoms is that each of these nuclear constituents are themselves composites built from still-smaller objects — the quarks and gluons of the previous chapter. Because of this substructure inter-nucleon forces are likely to be fairly complicated, in the same way that residual van der Waal's forces between atoms and molecules can appear more complicated than the Coulomb interaction despite it being the Coulomb interaction that is ultimately responsible for these forces. As we shall see, the analogy with residual inter-atomic forces in molecules can be a good one, and in many ways a nucleus is more similar to a very small drop of liquid than to an atom.

5.1 Nuclear binding energies and nucleon forces

Much of what is known about inter-nucleon forces comes from the observed properties of the nuclei which they combine together to make. This section summarizes some of those properties and uses them to argue that inter-nucleon forces are:

Short-ranged inasmuch as they act only over distances of order a few fm or less;

Attractive inasmuch as there are combinations of nucleons that can lower their energy by being in close proximity to one another, and thereby give rise to nuclei;

Democratic inasmuch as they are largely indifferent to nucleon spins or to the difference between a neutron or a proton (more precisely, they are invariant under the rotations and under isospin symmetries to be defined in §5.3 below);

Saturated inasmuch as nucleons interact with only a fixed number of neighbors in a way that is described in more detail in §5.1.3 below (and that makes nuclear matter incompressible).

5.1.1 Inter-nucleon interactions

Consider first what is implied by the statement that neutrons and protons experience a nuclear interaction that is strong, attractive and short-range.

The range of nuclear forces is around 1 fm, which is to say that nucleons separated by more than $\sim$ 1 fm largely do not experience any nuclear interactions. This range can be determined in many ways, such as by studying pp and pn scattering. For pp scattering purely Coulomb interactions stop giving the observed scattering cross section once impact parameters are roughly 1 fm or smaller. For np scattering the neutrality of the neutron ensures there is no Coulomb component, and the scattering cross section is very similar to a geometric cross section for an object whose radius is around 1 fm.

The inter-nucleon interaction must also have an attractive component because for some combinations of nucleons it gives rise to bound states (*i.e.* nuclei). Yet not all possible combinations of nucleons yield stable bound states, however, since (for example) there are no known nuclei consisting only of two or more neutrons. More concretely, nontrivial nuclei containing the fewest nucleons – *i.e.* nuclei with $A = 2$ – could in principle arise as pp, np or nn bound states, but of these only np actually appears in nature as a stable nucleus: the Deuterium nucleus or ^{2}H isotope of Hydrogen. The combinations pp and nn appear not to bind.

How strong is the inter-nucleon force? A good benchmark with which to compare it is the magnitude of the Coulomb repulsion of two protons that are separated by 1 fm, which is of order

$$U_C(1\text{ fm}) = \frac{\alpha}{1\text{ fm}} \simeq 1.4\text{ MeV}\,, \tag{5.1}$$

which uses $(1\text{ fm})^{-1} \sim 200$ MeV and where $\alpha \simeq 1/137$ is the fine-structure constant. Energies much larger than this are what we regard as 'strong' since they are capable of overwhelming the weaker Coulomb repulsion between protons.

To get an idea about the strength of the inter-nucleon force it is useful to try out a simple model for the neutron-proton attraction that gives rise to Deuterium. A simple model that captures the attractiveness, short range and strength of the p-n interaction is the attractive spherical square well sketched in Figure 27. This potential is characterized by two parameters: the force's range R and well's depth U_0.

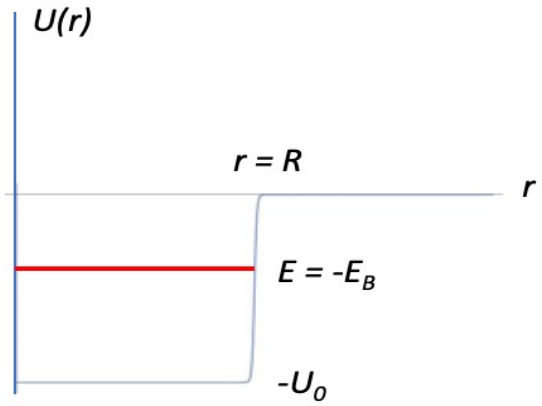


Figure 27. Sketch of the inter-nucleon interaction modelled as a square-well potential.

Bound states for this potential have energy $E = -E_B$, as shown in the figure. The magnitude of this bound-state energy is computed in §3.2.7 and are given by the joint solutions to the two equations (3.80) and (3.81), repeated here for convenience:

$$\kappa R = -(kR)\cot(kR) \quad \text{and} \quad (kR)^2 + (\kappa R)^2 = 2mU_0R^2\,, \tag{5.2}$$

where $\kappa^2 = 2mE_B$ and $k^2 = 2m(U_0 - E_B)$. The behaviour of these solutions can be visualized as the intersection of the two curves in Fig. 20, and leads to the result

$$E_B = \frac{\kappa_n^2}{2m} = \frac{c_n}{m_N R^2}\,, \tag{5.3}$$

where $m \simeq \frac{1}{2}\, m_N$ is the reduced mass of the two-nucleon system, with $m_N \simeq 940$ MeV being the *nucleon mass*. n is an integer and for each n the dimensionless quantity $c_n = c_n(mU_0R^2)$ is a monotonically increasing function of the parameter mU_0R^2.

As described in §3.2.7, the integer n arises because the number of independent bound-state solutions depends on the value of mU_0R^2. In particular, no bound states exist at all if $2mU_0R^2 < \frac{1}{2}\,\pi$, while there are $n = 1, 2, \cdots$ solutions whenever $(n - \frac{1}{2})\pi < 2mU_0R^2 < (n + \frac{1}{2})\pi$. If U_0 is dialled up with R fixed then new solutions arise with $E_B = 0$ and $kR = (n+\frac{1}{2})\pi$ whenever U_0 passes through values where $2mU_0R^2 = (n+\frac{1}{2})\pi$. Each new state then becomes more tightly bound (*i.e.* E_B grows) as U_0R^2 increases, eventually asymptoting (for large mU_0R^2) to $kR = (n+1)\pi$ and so

$$E_B = \frac{\kappa_n^2}{2m} \to U_0 - \frac{1}{2m}\left[\frac{(n+1)\pi}{R}\right]^2 \quad (\text{for } mU_0R^2 \gg 1)\,. \tag{5.4}$$

These results show that a key scale in the problem is the zero-point kinetic energy that is required by the uncertainty principle if a state is localized within a radius of size R. For such a localized state typical momenta are $k \sim 1/R$ and so (keeping in mind $2m \simeq m_N$) the associated zero-point kinetic energy is

$$E_{\rm zp} = \frac{k^2}{2m} \simeq \frac{1}{m_N R^2} \simeq 40 \text{ MeV} \left(\frac{1 \text{ fm}}{R}\right)^2 . \tag{5.5}$$

This energy scale is important because no bound states exist at all if U_0 is smaller than $E_{\rm zp}$, and new bound states arise every time U_0 increases by an amount of order $E_{\rm zp}$.

The strength of the inter-nucleon interaction can be roughly inferred by asking what depth of the potential would reproduce the observed Deuterium binding energy given a force of range $R \simeq 1$ fm. Since the measured Deuteron binding energy (the difference $m_n + m_p - m_D$ between the Deuteron mass and the mass of a free proton and neutron) is 2.2 MeV this implies the depth of the potential well must be

$$U_0 = \frac{k^2}{2m} + E_B \simeq E_{\rm zp} + E_B \simeq 42 \text{ MeV}\,, \tag{5.6}$$

where we use both (5.5) and $E_B \simeq 2$ MeV. This can easily overwhelm the Coulomb interaction of (5.1) acting over similar distances.

Since its binding energy is only 2.2 MeV the potential depth U_0 must be just slightly larger than $E_{\rm zp} \sim 40$ MeV, making the Deuteron a rather shallow bound state. Because it is so

shallowly bound its wave-function extends out to relatively large distances, $\psi(r) \propto (1/r)e^{-r/a}$ for $r > R$, with

$$a = \frac{1}{\kappa} = \frac{1}{\sqrt{2mE_B}} \sim \frac{1}{30\text{ MeV}} \sim 7\text{ fm} \gg R \sim 1\text{ fm}\,, \tag{5.7}$$

and so the Deuterium wave-function extends about seven times further than the assumed range of the nuclear force. Not only does the nuclear force have a definite range, this range can be short in comparison with the natural size of the particle wave-functions.

5.1.2 Short-range attractive forces and pairing

Short-range attractive forces have other implications, particularly when they act on identical particles (such as two protons or two neutrons). If these identical particles are fermions (as are nucleons) then short-ranged attraction can cause a preference for *pairing*: it is energetically preferable for pairs of particles to join together into spinless units.

To see why, imagine computing the energy eigenstates appropriate for a pair of particles that interact through a short-range attractive potential that depends only on their relative distance: $U = U(|\mathbf{r}_1 - \mathbf{r}_2|)$. Recalling the result from previous sections that the centre-of-mass motion can be separated from the relative motion, we label the two-particle states by its center-of-mass momentum, $\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2$ and total spin s, plus any labels (n, ℓ, ℓ_z) that describe the relative motion (which we choose to include the angular momentum quantum numbers $\ell = 0, 1, 2, \cdots$ and $\ell_z = -\ell, -\ell+1, \cdots, \ell-1, \ell$ because of the assumed rotational invariance of U:

$$|\mathbf{P}, s; n, \ell, \ell_z\rangle = \sum_{\sigma_1,\sigma_2=\pm\frac{1}{2}} \int \mathrm{d}^3\mathbf{r}_1 \mathrm{d}^3\mathbf{r}_2 |\mathbf{r}_1, \sigma_1; \mathbf{r}_2, \sigma_2\rangle \, \langle \mathbf{r}_1, \sigma_1; \mathbf{r}_2, \sigma_2 | \mathbf{P}, s; n, \ell, \ell_z\rangle\,, \tag{5.8}$$

where we choose a basis of position eigenstates for each of the individual particles.

Recall also that §3.2.1 shows that the two-particle wave-function factorizes if we use the pair's center-of-mass $\mathbf{R} = (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1 + m_2)$ and relative position $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ rather than $\mathbf{r}_1$ and $\mathbf{r}_2$ as position labels. That is,

$$\langle \mathbf{r}_1, \sigma_1; \mathbf{r}_2, \sigma_2 | \mathbf{P}, s; n, \ell, \ell_z\rangle = \frac{1}{(2\pi)^{3/2}} \exp(i\mathbf{P}\cdot\mathbf{R})\; \Psi_{n\ell\ell_z, s\sigma_1\sigma_2}(\mathbf{r})\,, \tag{5.9}$$

where Ψ satisfies the single-particle Schrödinger equation

$$-\frac{1}{2m}\nabla^2\Psi + U(r)\Psi = E\Psi\,, \tag{5.10}$$

where $m = m_1 m_2/(m_1 + m_2)$ is the reduced mass.

The preference for pairing into spinless combinations comes from the competition between Fermi statistic and the assumed short-range of the attractive force. On the one hand Fermi statistics states that the state must change sign if the two identical particles are interchanged, and so

$$\Psi_{n\ell\ell_z, s\sigma_2\sigma_1}(-\mathbf{r}) = -\Psi_{n\ell\ell_z, s\sigma_1\sigma_2}(\mathbf{r})\,. \tag{5.11}$$

As usual we exploit spherical symmetry to seek solutions of the form $\Psi(\mathbf{r}) = \mathcal{R}(r)Y_{\ell\ell_z}(\theta,\phi)$ in spherical polar coordinates. But in polar coordinates the reflection $\mathbf{r} \to -\mathbf{r}$ corresponds to $(r,\theta,\phi) \to (r,\pi-\theta,\phi+\pi)$, and the spherical harmonics satisfy the identity

$$Y_{\ell\ell_z}(\pi-\theta,\phi+\pi) = (-)^\ell Y_{\ell\ell_z}(\theta,\phi)\,, \tag{5.12}$$

so eq. (5.11) implies $\Psi_{n\ell\ell_z,s\sigma_2\sigma_1}(\mathbf{r}) = (-)^{\ell+1}\Psi_{n\ell\ell_z,s\sigma_1\sigma_2}(\mathbf{r})$. That is, states with ℓ even must be antisymmetric under the interchange $\sigma_1 \leftrightarrow \sigma_2$ (which — assuming spin-half particles, as for nucleons — implies the pair must have total spin $s=0$, since for this $2s+1=1$ corresponding to the unique antisymmetric spin combination $\uparrow\downarrow - \downarrow\uparrow$. Similarly states with ℓ odd must be symmetric under $\sigma_1 \leftrightarrow \sigma_2$, corresponding to total spin $s=1$ because there are $2s+1=3$ symmetric spin combinations: $\uparrow\uparrow$, $\downarrow\downarrow$ and $\uparrow\downarrow + \downarrow\uparrow$.

The short-range force comes in because it forces the lowest-energy state to have $\ell = 0$ and so therefore must also have its spins combine into the $s=0$ singlet state. Ground states very generally tend to be $\ell = 0$ because this minimizes the number of zeroes in the wavefunction, and every additional zero turns out to cost energy due to the need for larger gradients in the wavefunction. But this is clearest for short-range forces because for these any energy gain due to the interaction is concentrated very close to $r=0$, and for small r the radial equation (3.52) is given approximately by

$$\frac{\mathrm{d}^2\mathcal{R}}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}\mathcal{R}}{\mathrm{d}r} - \frac{\ell(\ell+1)}{r^2} \simeq 0\,, \tag{5.13}$$

which has solutions $\mathcal{R} = C_1 r^\ell + C_2 r^{-\ell-1}$. [Notice that this is consistent with the small-r limit of the explicit $\ell=0$ solutions of (3.77) once one recalls that $\mathcal{R}(r) = u(r)/r$.] Demanding Ψ not to diverge at $r \to 0$ then forces $C_2 = 0$ and so ensures $\Psi \propto r^\ell$ for small r. This shows that only $\ell = 0$ has the nonzero probability $|\Psi(\mathbf{r})|^2$ at $r=0$ that is required to profit from the attractiveness of a short-range potential that is localized at the origin.

The bottom line is this: because the inter-nucleon force is attractive and short-ranged we expect pairs of identical nucleons to prefer to minimize their energy by pairing together into spinless pairs. This should make nuclei with even numbers of protons and even numbers of neutrons (*i.e.* nuclei where A is divisible by 4 and Z is even) to be more tightly bound than are others. Although this simple pairwise argument might become suspect for nuclei containing many nucleons, where many-body effects might change the conclusion, it should be most reliable for nuclei with relatively few nucleons. This is the start of an explanation for why α particles – which are ^{4}He nuclei with two protons and two neutrons – seem to be extremely tightly bound.

5.1.3 Complex nuclei and saturation of nuclear forces

Ideas about the nature of inter-nucleon forces (such as the arguments for pairing given above) can be tested against the gross features of nuclei, and in particular by how their binding

energies depend on the number of neutrons and protons that are present. To this end it is useful to specify more precisely what is meant by nuclear binding energies.

Binding energies are defined by comparing the mass, $M(A,Z)$, of the isotope of interest with the sum of the masses of its constituents. Temporarily re-introducing factors of c:

$$-\frac{E_B}{c^2} = M(A,Z) - ZM(1,1) - (A-Z)M(1,0) = M(A,Z) - Zm_p - Nm_n\,. \qquad (5.14)$$

As defined E_B is positive when it costs energy to separate the isotope into its constituent protons and neutrons. Although in principle the masses appearing here should be of the nucleus only, in practice mass measurements are normally done with neutral atoms. This need not be a problem since the error (the difference in binding energy of the electrons) is much smaller than the nuclear binding energies.

In practice a more precise way to measure binding energies is by creating the isotope by bombarding another isotope with an appropriate beam (perhaps neutrons) and looking for the created nucleus to de-excite by emitting a photon. For example Deuterium (also called a deuteron) can be made by bombarding protons with neutrons through the reaction $n+p \to d+\gamma$. In this case the incoming neutron can have very low energy (a *thermal* neutron can have kinetic energies of order k_BT, which in lab settings are much smaller than an MeV. For such slow neutrons the photon then carries off the binding energy. Because it is possible to measure photon energies with much more precision than measuring isotope masses (which is typically done by following their trajectories in an applied magnetic field) binding energies can be more accurately known than are isotopic masses.

Symbol	E_B	E_B/A	ΔE	Symbol	E_B	E_B/A	Symbol	E_B	E_B/A
^{2}H	2.22	1.11	—	^{3}H	8.48	2.83	^{3}He	7.72	2.57
^{4}He	28.30	7.07	—	^{5}He	27.41	5.48	^{5}Li	26.33	5.27
^{6}Li	32.00	5.33	—	^{7}Li	39.25	5.61	^{7}Be	37.60	5.37
^{8}Be	56.50	7.06	−0.09	^{9}Be	58.17	6.46	^{9}B	56.31	6.26
^{10}B	64.75	6.48	—	^{11}B	76.21	6.93	^{11}C	73.44	6.68
^{12}C	92.16	7.68	7.27	^{13}C	97.11	7.47	^{13}N	94.11	7.24
^{14}N	104.66	7.48	—	^{15}N	115.49	7.70	^{15}O	111.96	7.46
^{16}O	127.62	7.98	14.44	^{17}O	131.76	7.75	^{17}F	128.22	7.54
^{18}F	137.37	7.63	—	^{19}F	147.80	7.78	^{19}Ne	143.78	7.57
^{20}Ne	160.65	8.03	19.17	^{21}Ne	167.41	7.97	^{21}Na	163.08	7.77
^{22}Na	174.15	7.92	—	^{23}Na	186.57	8.11	^{23}Mg	181.73	7.90
^{24}Mg	198.26	8.26	28.48	^{25}Mg	205.59	8.22	^{25}Al	200.53	8.02

Figure 28. Binding energies, E_B, and binding energy per nucleon, E_B/A, of the lightest stable nuclei in MeV. For nuclei with $A = 4n$ for $n = 2,3,4,5,6$ ΔE gives the binding energy relative to n times the binding of a ^{4}He nucleus. (Figure source: `Samuel Wong, ''Introductory Nuclear Physics,'' Wiley 2004.`)

Table 28 provides a table of the measured binding energies, E_B, for the lightest elements, as well as the binding energy per nucleon, E_B/A. Figure 29 provides a related plot of E_B/A versus A for a wider range of A. This table and the figure provide several lines of evidence in favour of the assertion that nuclei are held together because nucleons experience attractive short-range forces. To see why, first notice that for small nuclei the binding energy per nucleon grows as one adds more and more nucleons, as shown as the steep rise on the left of Figure 29. This occurs as more and more nucleon-nucleon pairs can profit from their mutual interaction to lower their energy.

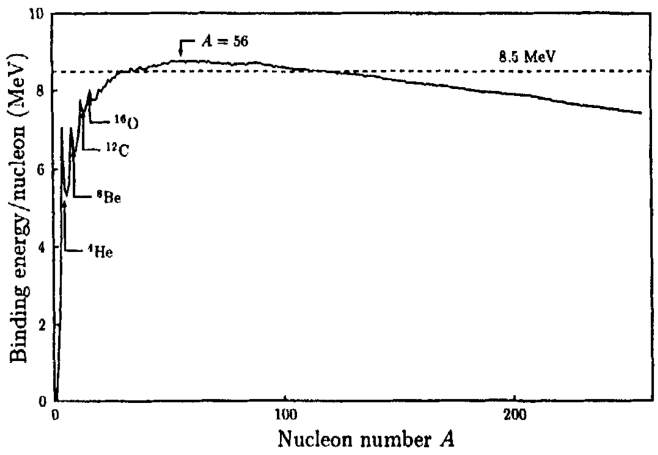


Figure 29. Plot of the binding energy per nucleon of nuclei vs the nucleus' total nucleon number A. (Figure source: `Samuel Wong, ''Introductory Nuclear Physics,'' Wiley 2004.`)

Both Table 28 and Figure 29 also show that this rise is erratic — at least for nuclei roughly out to ^{16}O — since light nuclei with $Z = N = 2n$ (so $A = 4n$ for n a small natural number) are more tightly bound than are others of similar size. This shows that the nuclear force is the strongest amongst the members of groups consisting of two neutrons and two protons, with additional nucleons not lowering energies in a commensurate way, and so is evidence for the nucleon pairing as discussed above. Apparently nucleons like to form α-particle clusters within nuclei — with $Z = 2$ and $A = 4$ — which then mutually interact and interact with any left-over unclustered nucleons.

The behaviour of Figure 29 for large A is also informative in a different way, because it provides evidence that the inter-nucleon force *saturates*. Here 'saturation' means that any particular nucleon only profits from the existence of a fixed number of nearby nucleons (rather than from all nucleons in a nucleus) when lowering its energy through its inter-nuclear interactions. Figure 29 shows evidence for saturation because the sharp growth of E_B eventually stops for intermediate-sized nuclei, after which the binding energy per nucleon

flattens out at around 8.5 MeV per nucleon. (The shallow downward trend in E_B/A for large A can be attributed to the repulsive nature of the Coulomb interaction between protons, which are more numerous for larger A – more about which below.)

Having E_B/A flatten for larger A implies total binding energy for large nuclei scales as $E_B \propto A$. The scaling $E_B \propto A$ shows that any one nucleon in a large nucleus is only interacting with a fixed number of other nucleons, so the binding energy per nucleon stops growing once there are more nucleons present than the maximum number of 'bonds' any one of them can form. Contrast this with the quadratic scaling of the Coulomb interaction for Z mutually interacting charges within a region of size R: $E_C \propto Z(Z-1)\alpha/R$. This quadratic dependence on the number of charges (Z) arises because the Coulomb interaction is long-ranged, and so every charge interacts with *all* the other charges. The factor $Z(Z-1)$ simply counts the number of such pairs of charges that can interact.

Other systems, like molecules in a fluid, also exhibit this kind of saturation and these also experience short-range forces (in this case often due to residual van der Waals interactions between atoms). But, in addition to having a short range, saturation of binding energies also usually requires the interaction to have a hard repulsive core in addition to the longer-range attractive component. The hard core stops the molecules from squeezing closer to one another to bring more of them within the range of the attractive short-range force, and thereby allow more pairs to profit from the lowered interaction energy. This suggests nucleons should also experience a short-range hard-core repulsion in addition to the longer-range attraction, and we find in later sections that such repulsion does exist because of the existence of 'exchange' interaction (described in more detail below).

As mentioned earlier, the gradual falloff of E_B/A seen for large A in Figure 29 can be attributed to the increased Coulomb repulsion due to the presence of more and more protons at larger A. But why have more protons at all? If nuclear forces are indifferent to whether they act on neutrons or protons why don't nuclei prefer to have more neutrons than protons? Indeed, why don't nuclei arise that are all neutrons, with no protons at all? This does not happen, as may be seen in Figure 30 which shows the stable (and relatively long-lived but radioactive) nuclei as a function of their neutron and proton number, N and Z. This reveals a relatively narrow *valley of stability*, with stable nuclei existing only for correlated values of N and Z. For small nuclei stability requires roughly equal numbers, $N = Z$, although it bends to allow $N > Z$ for larger nuclei (largely in response to the Coulomb penalty paid by having more protons).

The existence of this valley implies the existence of a nuclear *symmetry energy* that imposes a penalty for having N differ much from Z, and whose presence competes with the increased Coulomb energy cost of each proton. As we shall see this energy has its origin in the nucleon's Fermi statistics, which makes it energetically expensive to populate only neutron or only proton states.

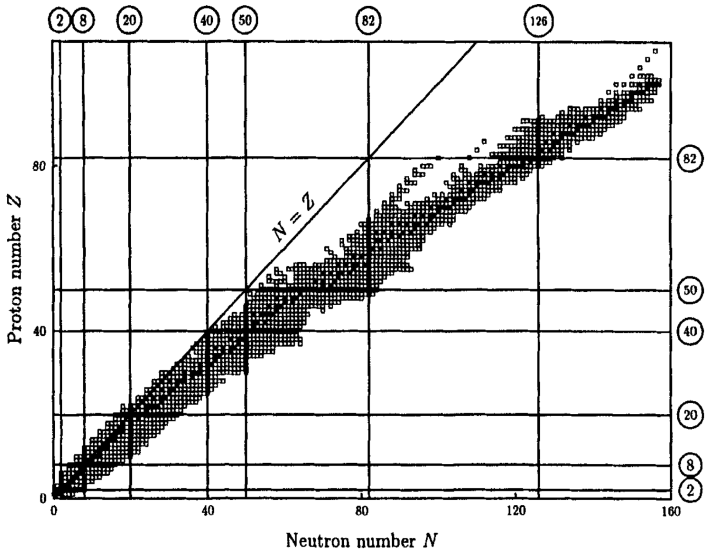


Figure 30. Plot of the 'valley of stability' showing the stable (black) and relatively long-lived radioactive (grey) isotopes vs proton (Z) and neutron (N) number. (Figure source: `Samuel Wong, ''Introductory Nuclear Physics,'' Wiley 2004.`)

5.2 Nuclear models

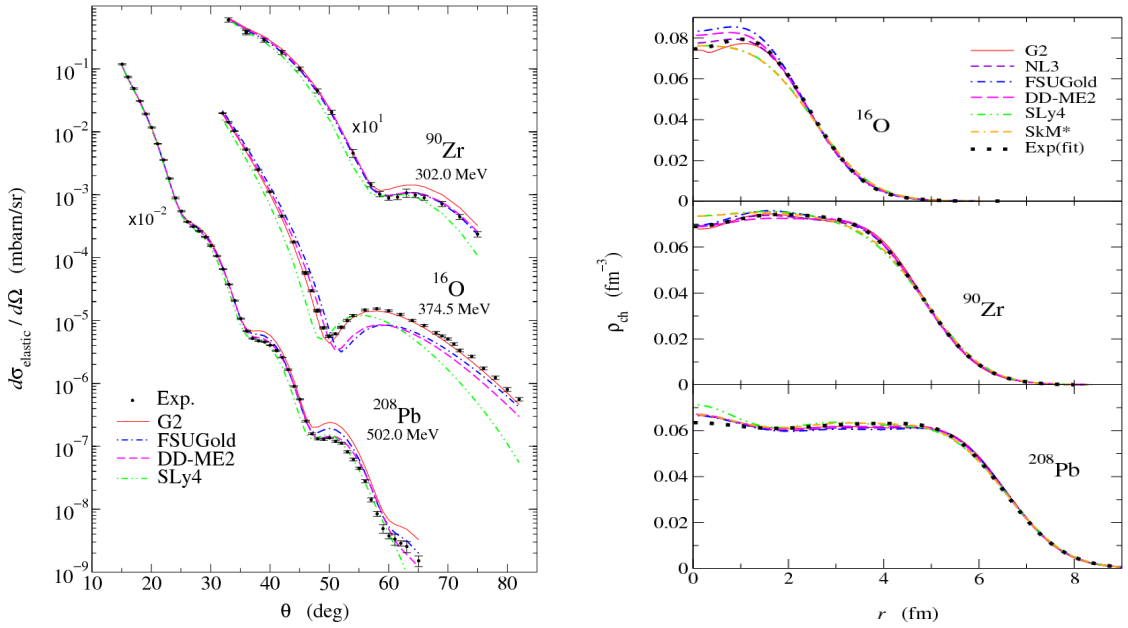
We now turn to several approximate ways to understand how nuclear interactions account for the observed properties of nuclei. The first of these asks how bulk properties like nuclear size and binding energy depend on the number of protons and neutrons, N and Z. The second studies more finely how quantum states for nucleons respond to this bulk environment.

5.2.1 Liquid-drop model

The evidence for nuclear saturation described above leads to a drop-like picture of the nucleus in which nucleons (like water molecules) interact dominantly with their immediate neighbours (attracting at longer distances with a repulsive hard core) and so form an incompressible fluid with a fixed energy density, whose volume is therefore proportional to the total number, A, of nucleons present. This picture is borne out by low-energy electron-scattering measurements of the nuclear density which (see Figure 31) show the diffractive peaks found earlier for hard-sphere scattering, indicating the existence of a nuclear surface.

For spherically symmetric nuclei the nuclear density that is inferred from experiments such as these are reasonably well-described by a *Fermi distribution* of the form

$$\rho(r) = \frac{N}{1 + \exp[(r - R)/a]}\,, \tag{5.15}$$



(a) Electron-nucleus scattering cross section

(b) Inferred nuclear charge distribution

Figure 31. Experimental determination of nuclear density from electron-nucleus scattering. Left panel: scattering cross section, showing diffractive peaks that indicate the presence of a nuclear surface. Right panel: inferred nuclear charge density obtained from scattering measurements (Figure source: `Phys.Rev.C78 (2008) 044332 arXiv:0808.1252 [nucl-th]`.)

where N is a normalization factor and R is called the *half-density radius*. The parameter a is related to the *surface thickness*, t, by $t = (4 \ln 3)a \simeq 4.39445\, a$, with t the distance over which ρ drops from 90% to 10% of its central value. Fits to these parameters give them a dependence on A that is approximately

$$R \simeq \left(1.18\, A^{1/3} - 0.48\right) \text{ fm} \qquad \text{and} \qquad t \simeq 2.4 \text{ fm}\,. \tag{5.16}$$

These values imply a central density of $\rho_0 = \rho(0) \simeq 0.17$ nucleons/fm^3. The scaling $R \propto A^{1/3}$ for large A confirms the expectation that the nuclear volume is proportional to A, as expected due to the saturation of nucleon interactions. By contrast, the surface thickness does not depend on nuclear size. Both of these are as would be expected for a droplet of fixed density. The shape implied by (5.15) is drawn in Figure 32, and has a clear resemblance to the measurements of Figure 31.

The liquid-drop picture also suggests a semi-empirical formula that captures the gross features of the dependence of the binding energy on N and Z, called the *Weizäcker mass formula*. This states

$$E_B(Z,N) = c_v\, A - c_s A^{2/3} - c_C \left[\frac{Z(Z-1)}{A^{1/3}}\right] - c_{\text{sym}} \left[\frac{(N-Z)^2}{A}\right] + \frac{\Delta}{A^{1/2}}\,, \tag{5.17}$$

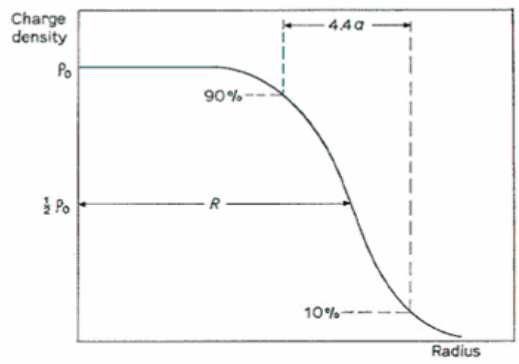


Figure 32. Plot of the Fermi nuclear charge distribution (which should be compared with the distribution inferred from measurements given in the right-hand panel of Figure 31). (Figure source: `http://202.141.40.218/wiki/index.php/Nuclear_Size_and_Radii`)

where

$$\Delta = \begin{cases} \delta & \text{for even-even nuclei} \\ 0 & \text{for odd-mass nuclei} \\ -\delta & \text{for odd-odd nuclei} \end{cases} , \tag{5.18}$$

and the constants are found by fitting to the observed binding energies and are given by

$$\begin{aligned} c_v &= 16 \text{ MeV} \qquad c_s = 17 \text{ MeV} \qquad c_C = 0.6 \text{ MeV} \\ c_{\text{sym}} &= 25 \text{ MeV} \qquad \delta = 12 \text{ MeV} . \end{aligned} \tag{5.19}$$

The terms in this expression have the following origin. The first (c_v or volume) term captures the amount by which the bulk energy density is reduced by the attraction of the inter-nucleon bonds. It contributes proportional to A because of the saturation of nuclear forces described above, which limits the number of such bonds for any one nucleon. As a result the total binding energy of this term simply counts the number of nucleons present.

The second (c_s or surface) term expresses how nucleons near the surface of the nucleus do not profit from the energy reduction of a full complement of bonds. This term scales like the surface area of the nucleus, and so varies like the square of the nuclear radius: R^2. Because the nuclear fluid has a fixed density the nuclear volume goes like $R^3 \propto A$, and so $R^2 \propto A^{2/3}$.

The third (c_C) term expresses the Coulomb repulsion of the protons, which we've seen scales like $Z(Z-1)$ and inversely with the nuclear radius, $R^{-1} \propto A^{-1/3}$.

The fourth (c_{sym} or symmetry) term arises due to the Fermi statistics of the nucleons (as we see below in more detail). The idea is that individual nucleons can in many ways be regarded as independently moving within a potential formed by the presence of all of the other nucleons. In this case each type of nucleon (protons and neutrons) will have a set of single-particle energy levels available, each of which Fermi statistics implies is occupied by

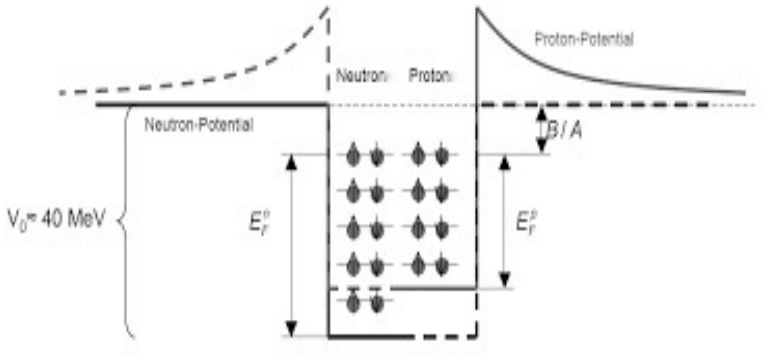


Figure 33. Sketch of the energy levels for independent protons (right) and neutrons (left) within a potential well. Notice the proton levels are displaced upwards relative to the neutrons by their average mutual Coulomb repulsion and their potential well also takes the Coulomb form outside the well. (Figure source: `http://web-docs.gsi.de/Lecture3.pdf`)

at most one particle (see Figure 33). In the ground state these levels are filled up to the point where all nucleons have been assigned a level, and so their total energy is dominated by the uppermost levels filled, called the *Fermi energy*, E_F. Since protons and neutrons have similar masses and interactions their Fermi energies are similar, and it is energetically punitive to just fill the nucleons with neutrons (thereby raising their Fermi energy) without equally filling the proton well. Consequently the sum of their Fermi energies is minimized when $N = Z$. Expanding the binding energy per nucleon in powers of $(N - Z)/A$ near this minimum therefore gives an energy $\delta E_B/A \propto (N-Z)^2/A^2$. It is this term whose competition with Coulomb repulsion determines how close to $N = Z$ are the stable nuclei.

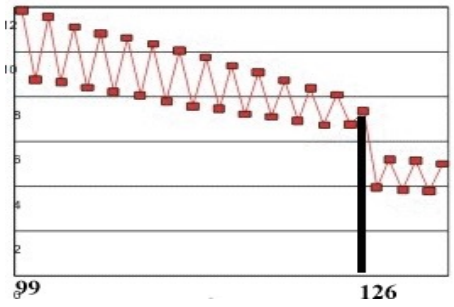


Figure 34. Plot of the energy cost (in MeV) of removing a neutron from Lead isotopes as a function of neutron number, N. The pattern of steps reflects the residual nucleon pairing interaction while the line at $N = 126$ corresponds to a nuclear *magic number* as described later in the text. (Figure source: `http://www.sjsu.edu/faculty/watkins/nnPb8201.gif`)

The final (Δ) term describes the effects of the nuclear pairing energy. Whenever a degen-

erate system of fermions experiences an attractive interaction it is energetically favourable for them to form pairs, since these pairs behave as bosons and so in many ways are released from the constraints of Fermi statistics. This phenomenon underlies a number of many-body effects involving fermions, such as superconductivity and superfluidity, and (because the nucleons are statistically degenerate) also contributes to the energetics of nuclei. There is considerable evidence for this pairing energy: (*i*) the larger number of stable or long-lived isotopes for even-even nuclei (for which both N and Z are even, so for which all nuclei can pair up) and the relative paucity of these states for odd-odd nuclei; (*ii*) even-even nuclei have spin zero in their ground states, indicating a preference for nucleon pairing into configurations with zero angular momentum; (*iii*) the energy cost for removing the outermost neutron (or *neutron separation energy* $S_n(A,Z) = E_B(A,Z) - E_B(A-1,Z)$) is systematically several MeV higher when N is even than when it is odd (see Figure 34) and a similar statement holds for the proton separation energy, $S_p(A,Z) = E_B(A,Z) - E_B(A-1,Z-1)$, when Z is even and odd.

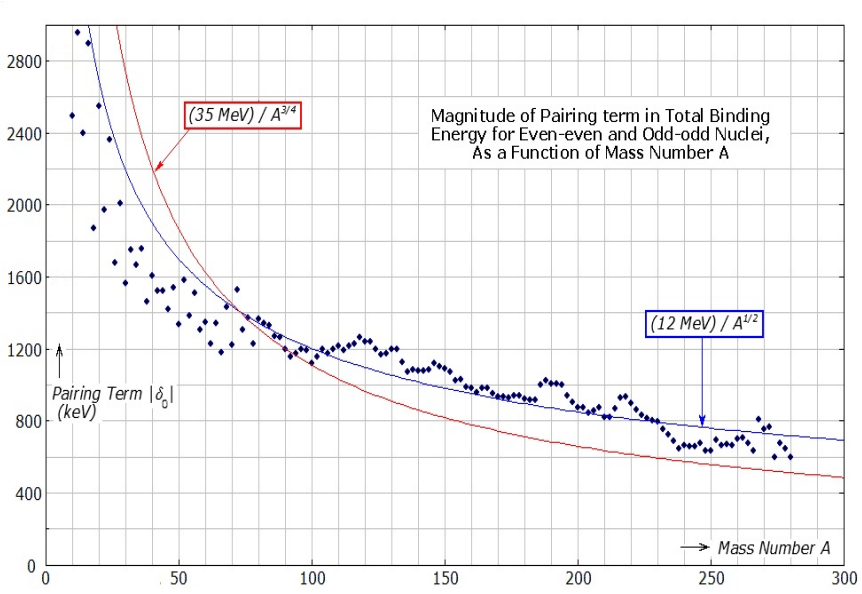


Figure 35. Example of a fit that leads to the parameters in the semi-empirical mass formula. This fit shows two fits to the A-dependence of the pairing energy). (Figure source: `Hans van Deukeren - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=51293587`

The Weizäcker formula expresses the main gross effects that contribute to nuclear binding energies and captures the competition between volume and surface (as well as Coulomb and symmetry) energies that shape the form of the nuclear valley of stability. For example Figure 36 plots $E_B(Z,A)$ as a function of Z for $A = 12$, showing how it is maximized for $Z = 6$ (corresponding to the stable nucleus ^{12}C). Table 8 evaluates the relative contributions of each contribution to E_B for each of the nuclei shown in this Figure, illustrating the trade-off (and

relative size) of each term. There are many more detailed features of binding energies that the Weizäcker formula does *not* capture, however, such as the existence of 'magic numbers' for which particularly stable nuclei exist. For these we turn to a slightly more refined model of nuclear structure.

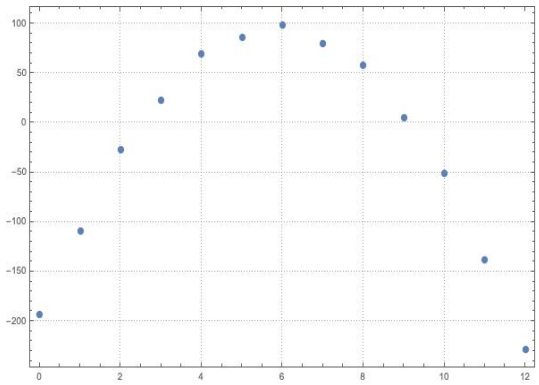


Figure 36. A plot of the binding energy $E_B(Z,A)$ vs Z for fixed $A = 12$, as computed using the semi-empirical mass formula (5.17) (with numerical coefficients from (5.19)). This is maximized at $Z_{\max} = 6$, corresponding to the stable nucleus ^{12}C. Such plots are useful for judging which types of radioactive decays are allowed energetically. For instance, instability towards the β decay $n \to p\, e^- \bar{\nu}$ occurs once $E_B(Z,A) < E_B(Z+1,A) - m_e$ (so *e.g.* ^{12}N$\to$ ^{12}C is allowed).

Table 8. Contributions (in MeV) to $E_B(Z,A)$ for $A = 12$

Element	Z	N	$E_{\rm vol}$	$E_{\rm surf}$	$E_{\rm em}$	$E_{\rm sym}$	$E_{\rm pair}$	E_B	E_B/A
Li	3	9	192	−89.1	−1.57	−75	−3.46	23	1.9
Be	4	8	192	−89.1	−3.14	−33.3	3.46	70	5.8
B	5	7	192	−89.1	−5.24	−8.33	−3.46	86	7.2
C	6	6	192	−89.1	−7.86	0.0	3.46	98	8.2
N	7	5	192	−89.1	−11.0	−8.33	−3.46	80	6.7
O	8	4	192	−89.1	−14.7	−33.3	3.46	58	4.9
F	9	3	192	−89.1	−18.9	−75	−3.46	5.6	0.46

Exercise 5.1: Use the Weizäcker mass formula to plot the binding energy of nuclei with $A = 127$ as a function of Z. What integer value of Z should be the most stably bound nucleus for this value of A? Evaluate the binding energy for the isobars $Z = 48, 49$ and $Z = 57, 58$. Based on this how would these four isobars decay? (*i.e.* do you expect them to dominantly experience α, β or γ decays, and if so to what should they decay?) What are the decay energies, Q, expected to

be released as kinetic energy by these decays? Later sections show that a rough estimate for the rate for β decay is $\Gamma \simeq G_F^2 Q^5/(4\pi)^3$, where Q is the energy released in the decay. If β decay is the dominant process, estimate the expected decay lifetimes for the above decays, using $G_F = 1.166379 \times 10^{-5}$ GeV^{-2}.

Exercise 5.2: Use the Weizäcker mass formula to compute the optimal (most tightly bound) value for Z for any given A. (Neglect the pairing energy when doing so.) Show that your result depends on both A and on $x := c_C/c_{\rm sym}$, and verify that $Z_{\rm opt} \to \frac{1}{2} A$ if $x \to 0$. Show that as $A \to \infty$ the optimal value instead satisfies $Z_{\rm opt} \to 2A^{1/3}/x$. Using the best-fit values for x what is the prediction for $Z_{\rm opt}$ when $A = 56$? How does this compare with the observed stable isotopes for this value of A?

5.2.2 Fermi gas model

The Fermi gas model of a nucleus considers each nucleon to move independently within a potential whose shape is meant to capture the average influence of all of the other nucleons. In the model the nucleons are considered to be independent of one another (*i.e.* non-interacting) since their interactions with the potential is meant to capture the main effects of nucleon-nucleon interactions. Because nucleons are fermions they cannot multiply occupy a quantum state and so they instead fill the available single-particle energy levels up to a nonzero Fermi energy, E_F, as in Figure 33.

At first sight this kind of picture might not be expected to be a very good description of a nucleus given the strong inter-nucleon interactions that are in play. It turns out not to do too badly, at least for heavier nuclei, for several reasons. One reason is that the residual inter-nucleon interactions can be taken to be somewhat weaker once the potential is included, since the potential already captures part of underlying inter-nucleon interactions (including, in particular, the binding of the nucleon to the nucleus). The other reason is due to the Fermi statistics satisfied by nucleons. At low energies most of the energy levels into which a nucleon might scatter are already occupied and so cannot be accessed because of Fermi statistics (this is often called 'Pauli blocking', in reference to the Pauli exclusion principle). This means that even a relatively strong interaction can have fairly weak matrix elements within a nuclear ground state, because the only nonzero matrix-elements connect a relatively narrow range of states very close to the Fermi energy. (This same argument is also why the free-electron model often works so well for conduction electrons in a metal.)

This picture leads to the simple understanding of the 'symmetry' term in the Weizäcker mass formula, on which we now elaborate in a bit more detail. The main observation is that the Fermi energy is determined by the number density of each type of nucleon. Consequently the Fermi energy for protons, $E_F(p)$, can be regarded as being a function of $Z/V \propto Z/A$, while the Fermi energy for neutrons, $E_F(n)$, is a function of $N/V \propto N/A$.

This is most explicit in the limiting case that the nucleons are regarded as free particles, whose single particle energy is $E(\mathbf{p}) = \mathbf{p}^2/2m_N$. Then, because momentum states are quantized for a particle living within a volume V, successively filling the least energetic momentum states with $\mathcal{N}$ particles fills them up to the Fermi momentum, p_F, defined by

$$\mathcal{N} = \int_0^{p_F} \mathrm{d}^3\mathbf{p} \left(\frac{\mathrm{d}n}{\mathrm{d}^3\mathbf{p}}\right) = 4\pi \int_0^{p_F} \mathrm{d}p\, p^2 \left(\frac{\mathrm{d}n}{\mathrm{d}^3\mathbf{p}}\right) , \tag{5.20}$$

where $\mathrm{d}n/\mathrm{d}^3\mathbf{p}$ is the particle's density of states (and is assumed in the second equality to be a function only of $p = |\mathbf{p}|$). So (for example) for a free particle (for which $\mathrm{d}n/\mathrm{d}^3\mathbf{p} = V/(2\pi)^3$, with V the system volume) p_F is given by

$$p_F^3 = 6\pi^2 \left(\frac{\mathcal{N}}{V}\right) = \frac{6\pi^2}{v_N} \left(\frac{\mathcal{N}}{A}\right) , \tag{5.21}$$

where the last equality uses $V \propto A$ and v_N is the average volume per particle. The Fermi energy then is $E_F = E(p_F)$, and so $2m_N E_F = p_F^2 = (6\pi^2\mathcal{N}/V)^{2/3}$, where $\mathcal{N} = N$ for neutrons and $\mathcal{N} = Z$ for protons.

The N and Z dependence of the symmetry energy near $Z = N$ does not depend on the detailed form of $E_F(p)$ and $E_F(n)$ as functions of N/A or Z/A, however. It may be found by expanding the Fermi energies about $Z = N$, by writing

$$N = \frac{A}{2}\left[1 + \frac{N-Z}{A}\right] \qquad \text{and} \qquad Z = \frac{A}{2}\left[1 - \frac{N-Z}{A}\right] , \tag{5.22}$$

so that, for instance, $E_F(N/A) = E_F(1/2) + E_F'(1/2)[(N-Z)/2A] + \cdots$, giving

$$\begin{aligned} E_{nF}(N/A) &\simeq E_{nF}^0 + \left[E_{nF}'\right]^0 \frac{N-Z}{2A} + \frac{1}{2}\left[E_{nF}''\right]^0 \left(\frac{N-Z}{2A}\right)^2 \\ \text{and} \quad E_{pF}(Z/A) &\simeq E_{pF}^0 - \left[E_{pF}'\right]^0 \frac{N-Z}{2A} + \frac{1}{2}\left[E_{pF}''\right]^0 \left(\frac{N-Z}{2A}\right)^2 , \end{aligned} \tag{5.23}$$

where the superscript '0' indicates evaluation at $Z/A = N/A = 1/2$. Consequently, assuming nuclear charge-independence implies $[E_{nF}']^0 = \left[E_{pF}'\right]^0$ the contribution of the Fermi energy to the nuclear binding budget is

$$E_{pF}(Z/A) + E_{nF}(N/A) \simeq \left[E_{pF}^0 + E_{nF}^0\right] + \frac{1}{8}\left[E_{pF}'' + E_{nF}''\right]^0 \left(\frac{N-Z}{A}\right)^2 . \tag{5.24}$$

Since this is a contribution to the binding energy per nucleon, multiplying through by A gives E_B. The first term is a contribution to the volume energy, c_v, while the second term is the leading contribution to the symmetry energy, c_{sym}, as advertised.

5.2.3 Shell model

The shell model elaborates on the Fermi gas picture by describing the nucleon energy levels in more detail. In particular, given the mean potential well the model also predicts the spacing and degeneracy of nucleon energy levels. This allows us to ask whether nuclei exhibit phenomena like 'closed shells' in the same way that closed electron shells do when explaining chemical properties in terms of atomic structure. There is evidence that nucleons also enjoy special properties near closed shells, which correspond to the existence of 'magic' values of N and Z where nuclei are particularly well bound (an example of which can be seen in Figure 34). These magic numbers are observed to occur when Z or N is equal to 2, 8, 20, 28, 50, 82 or 126.

To compute these energy levels and degeneracies requires first knowing the shape of the potential. Because of the short range and the saturation of nuclear forces the shape of the potential should be expected to track the average density of nucleons, such as given in (5.15) or Figure 32. This leads to a potential well whose shape — for spherical nuclei at least — lies somewhere between a spherical square well and a spherical harmonic oscillator. In both cases total angular momentum is conserved since the potentials are spherically symmetric, so states will be labelled by quantum numbers (n, ℓ, ℓ_z) with $\ell = 0, 1, 2, \cdots$ and $\ell_z = -\ell, -\ell + 1, \cdots, \ell - 1, \ell$, and n determined by solving the appropriate radial part of the Schrödinger equation.

Rotational invariance implies the energy eigenvalues found by solving the Schrödinger equation are independent of the magnetic quantum number ℓ_z: $E = E_{n\ell}$, so each energy level is degenerate by at least the $2(2\ell + 1)$ states corresponding to two spin states each for each choice of m. Consider first the harmonic oscillator potential,

$$V(r) = \frac{1}{2}\, M\, \omega^2\, r^2 = \frac{1}{2}\, M\, \omega^2 \left(x^2 + y^2 + z^2\right), \tag{5.25}$$

where the reduced mass $M = m_N$ is the nucleon mass (assuming the rest of the nucleus is much heavier than the nucleon of interest).

When solving for the energy levels we can either think of this as a radial harmonic oscillator for which ℓ, ℓ_z are the angular momentum quantum numbers, or we can think of this as three independent one-dimensional harmonic oscillators in the x, y and z directions. In this second way of thinking about it we know that energies are given by $E = \left(n_x + n_y + n_z + \frac{3}{2}\right)\hbar\omega$, and so the lowest energy state has $E_0 = \frac{3}{2}\,\hbar\omega$ and can be obtained only through the unique choice $n_x = n_y = n_z = 0$. It therefore must correspond to $\ell = 0$, and (keeping in mind the nucleon's two spin states) so has degeneracy 2. Following the nomenclature of atomic electron levels this ground state level is called the $1s$ orbital, where s corresponds to $\ell = 0$ and 1 is the value of the radial quantum number, n.

The next level has $E_1 = \frac{5}{2}\,\hbar\omega$ and can be obtained in one of three ways: by making $(n_x, n_y, n_z) = (1, 0, 0)$ or $(0, 1, 0)$ or $(0, 0, 1)$. This therefore corresponds to $\ell = 1$ and so has

degeneracy 6, and is called the $1p$ orbital. Next comes $E = \frac{7}{2}\,\hbar\omega$ which can be obtained from the following choices: $(n_x, n_y, n_z) = (2,0,0)$, $(0,2,0)$, $(0,0,2)$, $(1,1,0)$, $(1,0,1)$ and $(0,1,1)$. These six combinations corresponds to the $2(2\ell+1) = 10$ states appropriate to $\ell = 2$ of the $1d$ level plus the $2(2\ell+1) = 2$ states of the $2s$ level, and so on.

Table 9. Nucleon shells for the 3D Harmonic Oscillator

N	Orbital	Parity	Degeneracy	Cumulative number of levels
0	$1s$	$+$	2	2
1	$1p$	$-$	6	8
2	$2s$, $1d$	$+$	12	20
3	$2p$, $1f$	$-$	20	40
4	$3s$, $2d$, $1g$	$+$	30	70
5	$3p$, $2f$, $1h$	$-$	42	112
6	$4s$, $3d$, $2g$, $1i$	$+$	56	168

A list of the degeneracies of several more of the lowest harmonic-oscillator levels are given in Table 9, labelled by their value of $n_{\rm tot} = n_x + n_y + n_z$ and their *parity* (*i.e.* how they behave under spatial inversion: $\psi(-\mathbf{r}) = \pm\psi(\mathbf{r})$). The parity of a state in the shell model is dictated by its orbital angular momentum, and since spherical harmonics satisfy $Y_{\ell\ell_z}(\theta,\phi) = (-)^\ell Y_{\ell\ell_z}(\pi-\theta,\phi+\pi)$ their parity is simply $(-)^\ell$. This means all s, d, g and i orbitals are parity even while p, f and h are parity odd. Notice that the degeneracies of these first few energy levels of Table 9 precisely reproduce the first few magic numbers: 2, $2+6=8$ and $2+6+12=20$, but not the remaining ones.

Part of the problem with getting the remaining magic numbers is specific to the harmonic oscillator potential, since it is not generic for rotationally invariant potentials that, for instance, the $2s$ and $1d$ levels have the same energy. This is not true of the square-well potential, for example, and so a more realistic potential shape is expected to lift these harmonic-oscillator specific degeneracies. This is indeed what happens, and because the flatter bottom of the realistic potential deepens the potential for larger r relative to the harmonic oscillator, it has the effect of lowering the energy of the larger-ℓ states (thereby splitting, for instance, the $2s$ and $1d$ states by lowering $1d$ relative to $2s$). This leads to the more accurate ordering of orbitals shown in Table 10. Notice the splittings can be large enough to mix up the harmonic oscillator levels: the $n_{\rm tot} = 5$ $1h$ state moves down below the uppermost ($3s$) $n_{\rm tot} = 4$ state. Similarly, $1i$ of the $n_{\rm tot} = 6$ level gets depressed relative to the $3p$ level of the $n_{\rm tot} = 5$ harmonic oscillator level.

We see from the table that the larger magic numbers at 50, 82 and 126 remain (so far) unexplained. The missing ingredient is a spin-dependent interaction, $\delta V_{so} = -C_{so}\,\mathbf{L}\cdot\mathbf{s}$,

Table 10. Nucleon shells for more realistic potentials

Initial level number	Orbital	Parity	Degeneracy	Cumulative number of levels
0	$1s$	+	2	2
1	$1p$	−	6	8
2	$1d$	+	10	18
2	$2s$	+	2	20
3	$1f$	−	14	34
3	$2p$	−	6	40
4	$1g$	+	18	58
4	$2d$	+	10	68
5	$1h$	−	22	90
4	$3s$	+	2	92
5	$2f$	−	14	106
6	$1i$	+	26	132
5	$3p$	−	6	138

where C_{so} is a constant while $\mathbf{L}$ is the nucleon's orbital angular momentum and $\mathbf{s}$ is its spin. This type of interaction splits the degeneracy of the $2(2\ell+1)$ states in each orbital because the different nucleon spin states can now have different energies. Furthermore, because the interaction is proportional to $\mathbf{L}$ its effects are biggest for the largest ℓ and so change more the f, g and h orbitals than the s, p and d ones. As a result this kind of interaction can alter the large magic numbers without changing much the smaller ones. Spin-dependent couplings also play a role in atomic electron levels, through the spin-orbit interaction.

But are nucleon-nucleon interactions really spin-dependent? There is good evidence they are, an example of which is given by the properties of the ground and first excited states of the isotope ${}^{5}_{2}$He. We know ${}^{4}_{2}$He is very tightly bound and spinless, as expected for a 'doubly magic' nucleus where both Z and N take the lowest magic number corresponding to filling the $1s$ orbital for both protons and neutrons. We expect the additional neutron in ${}^{5}_{2}$He to be in the $1p$ orbital, but because of the neutron spin the rules of combining angular momenta tell us this $\ell = 1$ orbital can have total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{s}$ with quantum number $j = \ell + \frac{1}{2} = \frac{3}{2}$ or $j = \ell - \frac{1}{2} = \frac{1}{2}$. What is found experimentally is that the $j = \frac{3}{2}$ combination has slightly lower energy than the $j = \frac{1}{2}$ combination, so the ${}^{5}_{2}$He ground state has $j = \frac{3}{2}$ and its first-excited state has $j = \frac{1}{2}$. Evidently some sort of inter-nucleon interaction is depressing the $j = \frac{3}{2}$ elements of the $1p$ orbital relative to the $j = \frac{1}{2}$ ones.

The interaction $V_{so} = -C_{so}\,\mathbf{L}\cdot\mathbf{s}$ can do precisely this. In particular, using $\mathbf{J}^2 = (\mathbf{L}+\mathbf{s})^2 = \mathbf{L}^2 + \mathbf{s}^2 + 2\,\mathbf{L}\cdot\mathbf{s}$ and that the eigenvalues of $\mathbf{J}^2$, $\mathbf{L}^2$ and $\mathbf{s}^2$ are respectively given by $j(j+1)$,

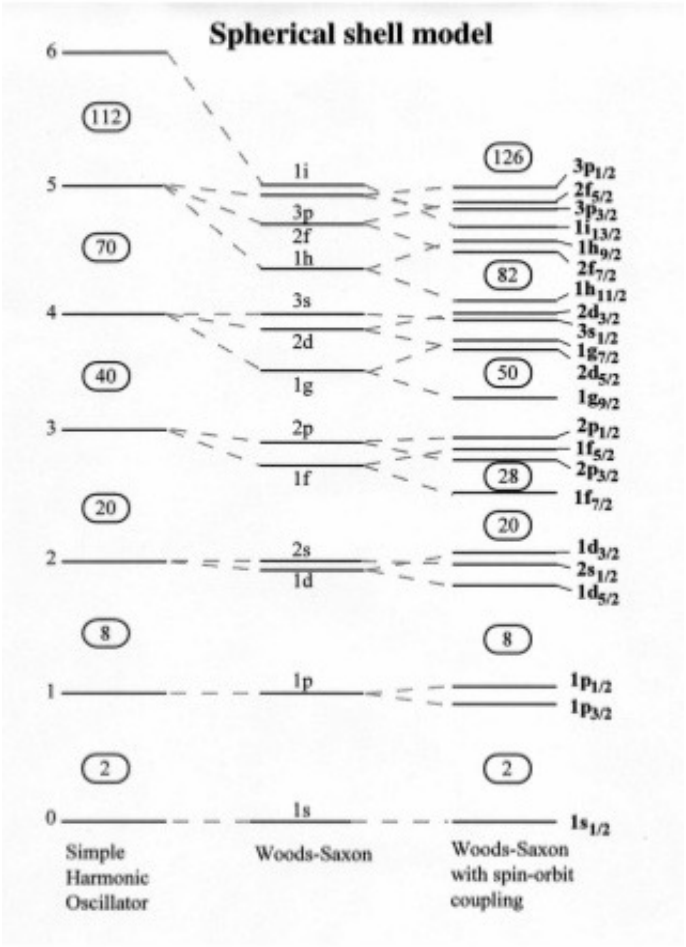


Figure 37. Diagram showing how harmonic oscillator energy levels split when replaced by a more accurate potential with a flatter bottom and then split again once spin-dependent interactions are added. The final level diagram does reproduce the observed magic numbers. (Figure source: `http://www.slideshare.net/brucelee55/nuclear-isomerism-probes-of-nuclear-structure`)

$\ell(\ell+1)$ and $s(s+1)$, the eigenvalues of V_{so} acting on a state with quantum numbers $|j,\ell,s\rangle$ are

$$E_{so}(j,\ell,s) = -\frac{1}{2}\,C_{so}\left[j(j+1) - \ell(\ell+1) - s(s+1)\right]. \tag{5.26}$$

For nucleons we have $s = \frac{1}{2}$ and $j = \ell \pm \frac{1}{2}$ and so this becomes

$$E_{so}(j=\ell+1/2) = -\frac{1}{2}\,C_{so}\,\ell \qquad \text{and} \qquad E_{so}(j=\ell-1/2) = \frac{1}{2}\,C_{so}\,(\ell+1)\,, \tag{5.27}$$

which gives a splitting of $\Delta E_{so} = -C_{so}\,\left(\ell+\frac{1}{2}\right)$, whose magnitude grows with ℓ (as expected). Notice that for the particular case of the $1p$ state (as appropriate for $^{5}_{2}$He) the initially 6-fold

degenerate $1p$ state becomes split into $1p_{1/2}$ and $1p_{3/2}$ states (where the subscript gives the value taken for j), and the $1p_{3/2}$ state has lower energy if $C_{so} > 0$.

This same interaction also allows the level diagrams described earlier to account for the larger magic numbers (see Figure 37). The key point is that the splitting of the largest-ℓ levels (*i.e.* $1g$, $1h$ and $1i$ levels) is big enough that it pushes the $j = \ell + \frac{1}{2}$ state down into the lower shell but keeps the $j = \ell - \frac{1}{2}$ state from also doing so. For example, Figure 37 shows that the first level for which this happens is the $1g_{9/2}$ level, which drops down to the next lowest shell and so adds its $2j + 1 = 10$ states to the 40 already in these lower shells to achieve the magic number 50. Similarly, only the $1h_{11/2}$ state drops down to the shell below with its $2j + 1 = 12$ states, leaving the 10 states of the $1h_{9/2}$ orbital in the higher shell. This then properly reproduces the magic number at 82. Finally, only the 14 states of the $1i_{13/2}$ orbital actually move to the lower shell, leaving the 12 of $1i_{11/2}$ in the upper level. This gives the magic number at 126, and so on.

Table 11. Nuclei with closed shells plus an additional nucleon

Nuclide	Z	N	Shell-model prediction	Observed spin/parity
^{17}O	8	9	$d_{5/2}$	$\frac{5}{2}^+$
^{17}F	9	8	$d_{5/2}$	$\frac{5}{2}^+$
^{41}Sc	21	20	$f_{7/2}$	$\frac{7}{2}^-$
^{209}Pb	82	127	$g_{9/2}$	$\frac{9}{2}^+$
^{209}Bi	83	126	$h_{9/2}$	$\frac{9}{2}^-$

Similar to what happens for the shell model of atomic electrons, the shell model for nucleons predicts the properties of all nuclei that are obtained from closed shells by adding or removing a single nucleon. In particular, since the closed shell is spinless the spin of the ground state with one extra (or one missing) nucleon is given by the j for the nucleon in question. The shell model does reasonably well on this score, properly predicting the properties of these nuclei (Table 11 shows several examples).

Exercise 5.3: The spin-orbit Hamiltonian for a particle of mass m and radial distance r from the axis of rotation has the form

$$H_{so} = -\frac{\mathbf{S} \cdot \mathbf{L}}{mr^2} \,.$$

Use this expression and the properties of inter-nucleon forces discussed in §5.2.1 to estimate how the constant C_{so} in eq. (5.26) depends on the atomic weight A. Find the value of C_{so} required to account for the 4 MeV splitting between the $j = \frac{3}{2}$ ground state of ^{5}He and its lowest $j = \frac{1}{2}$ excited state.

Exercise 5.4: Use the single-particle shell model to explain why islands of isomerism exist. (Long-lived excited nuclear states are called isomers, and 'islands of isomerism' refers to the fact that isomers arise for groups of nuclei that all have similar values for Z and A.) In particular, explain why the excited state with energy 0.225 MeV of the nuclide ^{85}Sr has a fairly long half-life (which turns out to be about 70 minutes). [HINT: Possibly useful for this explanation is the observation that emission of a photon (the usual way a nuclear state de-excites) changes nuclear angular momentum by $\Delta j = \pm 1$ since the photon has spin one.]

5.3 Isospin and meson exchange

The inter-nucleon potential turns out to be relatively complicated in its details, as might be expected of a residual interaction between pairs of objects that are themselves built from smaller things. The complications of this potential can be inferred most easily from the features of Deuterium, which we've seen most pristinely involves its implications only for a single pair of nucleons. This section describes some of these complications, leading to the concept of isospin and of interactions having their origin (at least in part) due to the exchange of pions.

We start by formulating more precisely the spin-dependence of the inter-nucleon forces, since this sets up the language with which to treat the charge-independence of the nucleon-nucleon interaction.

There are really several nuclear potentials under discussion when talking about inter-nucleon interactions, depending on whether or not we have protons interacting with protons, protons with neutrons or neutrons with neutrons. For scattering our interest really is with the matrix elements of these interactions between initial and final states, for instance in the Born approximation we seek the Fourier transform of the potential, which has the form

$$\langle f|V|i\rangle = \int \mathrm{d}^3\mathbf{x}\; \psi_f^*(\mathbf{x})\, V(\mathbf{x})\, \psi_i(\mathbf{x}) \propto \int \mathrm{d}^3\mathbf{x}\; V(\mathbf{x})\; \exp\Big[i(\mathbf{p}_i - \mathbf{p}_f)\cdot\mathbf{x}\Big]\,, \tag{5.28}$$

where the initial and final states are plane waves, $\psi_i(\mathbf{x}) \propto \exp\big[i\mathbf{p}_i{\cdot}\mathbf{x}\big]$ and $\psi_f(\mathbf{x}) \propto \exp\big[i\mathbf{p}_f{\cdot}\mathbf{x}\big]$, that describe the *relative* motion of the scattering particles once their overall centre-of-mass motion is factored out.

5.3.1 Spin-dependent and tensor interactions

In the previous sections we saw that the inter-nucleon force is spin-dependent; how is this incorporated in the above expression? Although spin-dependence is not in itself pertinent to the issue of charge-independence of nuclear forces, it is worth digressing to discuss how to incorporate it since the tools used do play a role for charge-independence. It is tempting simply to say that both of ψ_i and ψ_f should be given a two-component spin label: $\psi_{i\uparrow}$ and $\psi_{i\downarrow}$ and similarly for $\psi_{f\uparrow}$, $\psi_{f\downarrow}$, and so promote V into a two-by-two matrix of potentials: $V_{\uparrow\uparrow}$,

$V_{\uparrow\downarrow}$, $V_{\downarrow\uparrow}$ and $V_{\downarrow\downarrow}$. However this is not general enough since in scattering there are *two* initial and final nucleons, and so their combined spins can take any of $2 \times 2 = 4$ initial and final configurations; the scattering matrix should really be a 4×4 matrix rather than just a 2×2 one. We are being fooled by the form of (5.28), which for the position label gives the illusion of a one-particle problem because of our removal of the centre-of-mass motion.

The proper formulation of spin-dependent two-body interactions therefore instead thinks of $\langle f|V|i\rangle$ as a 4-by-4 matrix in spin space, with both $|i\rangle$ and $\langle f|$ described by *two* spin labels (one each for the spin of the initial or final particles): *e.g.* $|i\rangle = |\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle$ and similarly for $\langle f|$. The general spin-dependent potential then becomes a collection of 16 possible combinations: $\langle a, b|V(\mathbf{x})|c, d\rangle = V_{ab;cd}(\mathbf{x})$, where each of the labels a, b, c and d takes the two values $\uparrow$ and $\downarrow$.

This all sounds fairly complicated, but we also know that interactions should be rotation-invariant and this limits the kinds of matrices that can appear in $V_{ab;cd}(\mathbf{x})$. Rotation invariance states that $V_{ab;cd}(\mathbf{x})$ must be invariant under arbitrary 2-by-2 unitary rotations of $|i\rangle$ and $|f\rangle$ that mix up $\uparrow$ and $\downarrow$ simultaneously for both particles:

$$\begin{pmatrix} |\uparrow\uparrow\rangle \\ |\downarrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\downarrow\rangle \end{pmatrix} \to \mathcal{U} \begin{pmatrix} |\uparrow\uparrow\rangle \\ |\downarrow\uparrow\rangle \\ |\uparrow\downarrow\rangle \\ |\downarrow\downarrow\rangle \end{pmatrix} \qquad \text{and} \qquad \Big(\langle\uparrow\uparrow| \ \langle\downarrow\uparrow| \ \langle\uparrow\downarrow| \ \langle\downarrow\downarrow|\Big) \to \Big(\langle\uparrow\uparrow| \ \langle\downarrow\uparrow| \ \langle\uparrow\downarrow| \ \langle\downarrow\downarrow|\Big)\,\mathcal{U}^\dagger\,. \quad (5.29)$$

In practice this means V must be built only from the unit matrix and dot products of some combination of the spin matrices for each particle, $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$, with each other or with another vector, such as the relative angular momentum, $\mathbf{L}$, or the relative position, $\mathbf{x}$. Here the spin matrices $\mathbf{s}^{(1)} = \frac{1}{2}\boldsymbol{\sigma} \otimes I$ and $\mathbf{s}^{(2)} = I \otimes \frac{1}{2}\boldsymbol{\sigma}$ act separately on each of the nucleon spins:

$$\langle f|\,\mathbf{s}^{(1)}|i\rangle = \sum_{abcd}\langle ab|\,\mathbf{s}^{(1)}_{ab;cd}\,|cd\rangle = \frac{1}{2}\sum_{abcd}\langle ab|\,\boldsymbol{\sigma}_{ac}\,\delta_{bd}\,|cd\rangle\,, \quad (5.30)$$

and

$$\langle f|\,\mathbf{s}^{(2)}|i\rangle = \sum_{abcd}\langle ab|\,\mathbf{s}^{(2)}_{ab;cd}\,|cd\rangle = \frac{1}{2}\sum_{abcd}\langle ab|\,\delta_{ac}\,\boldsymbol{\sigma}_{bd}\,|cd\rangle\,. \quad (5.31)$$

Here $\boldsymbol{\sigma}$ denotes the usual Pauli matrices: *i.e.* the vector of matrices: $\boldsymbol{\sigma} = \{\sigma^1, \sigma^2, \sigma^3\}$, with

$$\sigma^1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \sigma^3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (5.32)$$

Notice that these definitions do what is expected for spin-half objects. In particular, because the Pauli matrices all square to the 2-by-2 unit matrix, we have $\mathbf{s}^{(1)}\cdot\mathbf{s}^{(1)} = \mathbf{s}^{(2)}\cdot\mathbf{s}^{(2)} = \frac{3}{4}\mathcal{I}$, where $\mathcal{I}$ is the 4-by-4 unit matrix (as opposed to the 2-by-2 unit matrix I), in agreement with the expectation that these should be $s(s+1)\mathcal{I}$ for the case $s = \frac{1}{2}$. Similarly the sum $\mathbf{s} = \mathbf{s}^{(1)} + \mathbf{s}^{(2)}$ squares to give a 4-by-4 matrix $\mathbf{s}\cdot\mathbf{s}$ whose four eigenvalues are $\{\lambda_i\} = \{2, 2, 2, 0\}$.

These correspond to $\mathbf{s}\cdot\mathbf{s}$ having eigenvalue $s(s+1)=2$ for a 3-by-3 space of states with combined total spin 1 (spanned by the states $|\uparrow\uparrow\rangle$, $|\downarrow\downarrow\rangle$ and $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle+|\downarrow\uparrow\rangle)$) as well as having eigenvalue zero when acting on the state $\frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle-|\downarrow\uparrow\rangle)$ (which has total spin 0).

Exercise 5.5: In the 4-dimensional basis of states given by (5.29) show that the definitions (5.30), (5.31) and (5.32) imply $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$ are given by the following explicit 4-by-4 matrices:

$$s_x^{(1)}=\frac{1}{2}\begin{pmatrix}\sigma_x & 0\\ 0 & \sigma_x\end{pmatrix}\,,\quad s_y^{(1)}=\frac{1}{2}\begin{pmatrix}\sigma_y & 0\\ 0 & \sigma_y\end{pmatrix}\,,\quad s_z^{(1)}=\frac{1}{2}\begin{pmatrix}\sigma_z & 0\\ 0 & \sigma_z\end{pmatrix}\,,$$
$$s_x^{(2)}=\frac{1}{2}\begin{pmatrix}0 & I\\ I & 0\end{pmatrix}\,,\quad s_y^{(2)}=\frac{1}{2}\begin{pmatrix}0 & -iI\\ iI & 0\end{pmatrix}\,,\quad s_z^{(2)}=\frac{1}{2}\begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix}\,,$$

where each entry is itself a 2-by-2 matrix (with I denoting the 2-by-2 unit matrix). Use these expressions to verify $\mathbf{s}^{(1)}\cdot\mathbf{s}^{(1)}=\mathbf{s}^{(2)}\cdot\mathbf{s}^{(2)}=\frac{3}{4}\mathcal{I}$ and that the eigenvalues of the matrix $\mathbf{s}\cdot\mathbf{s}$ – with $\mathbf{s}:=\mathbf{s}^{(1)}+\mathbf{s}^{(2)}$ – are $\{2,2,2,0\}$.

Examples of spin-dependent rotationally invariant interactions written this way then are

$$V(\mathbf{r})=V_{LS}(r)\;\mathbf{L}\cdot\left[\mathbf{s}^{(1)}+\mathbf{s}^{(2)}\right]+V_{SS}(r)\;\mathbf{s}^{(1)}\cdot\mathbf{s}^{(2)}+V_T(r)\;S_{12}\,, \tag{5.33}$$

and so on. Only a few of the many possible terms are written explicitly here. The first of these is essentially a spin-orbit coupling along the lines discussed above for the shell model, while the second represents a spin-spin interaction. The last term is called a *tensor interaction*, with the quantity S_{12} defined by

$$S_{12}=3\,(\mathbf{s}^{(1)}\cdot\mathbf{e}_r)(\mathbf{s}^{(2)}\cdot\mathbf{e}_r)-\mathbf{s}^{(1)}\cdot\mathbf{s}^{(2)}\,, \tag{5.34}$$

and its explicit dependence on the direction of $\mathbf{r}$, implied by the appearance of $\mathbf{e}_r:=\mathbf{r}/r$, ruins the invariance under rotations of $\mathbf{r}$ — with the $\mathbf{s}^{(i)}$ fixed — on which the conservation of orbital angular momentum, $\mathbf{L}$, relies.[28]

The properties of Deuterium provide several lines of evidence for the existence of a tensor interaction like V_T in the inter-nucleon force. One such involves the deuteron magnetic moment, which does not come out quite right when combining just the magnetic moments of its constituent nucleons for the $\ell=0$ ground state that arises when angular momentum is conserved. The observed value turns out to be consistent with expectations if the deuteron actually contains a small admixture of a d-wave $\ell=2$ state in addition to the main, s-wave $\ell=0$, component. But this kind of mixture of two different angular momenta is only possible if the potential depends on the direction of $\mathbf{x}$ in addition to its magnitude (as does the tensor interaction).

[28] Notice that V_T remains rotation invariant provided the spins, $\mathbf{s}^{(i)}$, rotate in the same way as does $\mathbf{r}$.

Additional evidence for the admixture of $\ell = 2$ states in the deuteron (and so also for the tensor interaction) comes from the fact that the deuteron is measured to have a nonzero electric quadrupole moment. This is forbidden if the deuteron only consists of $\ell < 2$ states, but is also explained if there is a small $\ell = 2$ component to the ground state.

5.3.2 Isospin and charge-independence of nucleon forces

With these tools in hand we can now state more precisely what is meant by the charge-independence of the inter-nucleon force. The idea is to consider the proton and neutron as if they are two 'spin-like' components of the nucleon: $N_\uparrow = p$ and $N_\downarrow = n$. This is not meant as anything to do with real spin, but is instead a useful way to think about the neutron and proton label that distinguishes the two types of nucleon. In this way we can write the four single-nucleon spin states as $|t, s\rangle$, where the spin quantum number is $s = \uparrow, \downarrow$ (and labels the 3rd component of spin as usual) while the *isospin* (or *isotopic spin*) label $t = \uparrow$ corresponds to a proton while $t = \downarrow$ corresponds to a neutron. That is

$$\begin{aligned} |p(s=\uparrow)\rangle &= |t=\uparrow, s=\uparrow\rangle\,, \quad |p(s=\downarrow)\rangle = |t=\uparrow, s=\downarrow\rangle\,, \\ |n(s=\uparrow)\rangle &= |t=\downarrow, s=\uparrow\rangle\,, \quad |n(s=\downarrow)\rangle = |t=\downarrow, s=\downarrow\rangle\,. \end{aligned} \tag{5.35}$$

In this language the charge-independence of nuclear forces can be phrased as the requirement that the interactions are invariant under arbitrary 2×2 unitary rotations acting on the nucleon isospin index. That is, it is invariant under an 'isospin rotation'

$$\begin{bmatrix} |p(s)\rangle \\ |n(s)\rangle \end{bmatrix} \to U \begin{bmatrix} |p(s)\rangle \\ |n(s)\rangle \end{bmatrix}\,, \tag{5.36}$$

where $UU^\dagger = U^\dagger U = I$, with I representing the 2×2 unit matrix. Any such a matrix is described by four complex entries subject to the four real conditions implied by $U^\dagger U = I$, and so can be written in terms of four real free parameters.

A convenient choice for these parameters is

$$U = e^{i\theta}\, \exp\left[\frac{i}{2}\,\vec{\omega}\cdot\vec{\tau}\right]\,, \tag{5.37}$$

where the four real parameters are θ and $\vec{\omega} = \{\omega_1, \omega_2, \omega_3\}$ and $\vec{\omega}\cdot\vec{\tau} = \omega_1\tau_1 + \omega_2\tau_2 + \omega_3\tau_3$, where τ_a again denote the three Pauli matrices

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,, \qquad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{and} \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,. \tag{5.38}$$

Although Pauli matrices are usually defined acting on the spin degrees of freedom, s, s', these instead act on the 'proton-neutron' degree of freedom. (In what follows we reserve the symbol $\boldsymbol{\sigma}$ for the vector of Pauli matrices acting on the spin degrees of freedom, and $\vec{\tau}$ for the isovector of Pauli matrices acting on the isospin label, t.)

The vector symbol (or arrow superscript) for $\vec{\omega}$ and $\vec{\tau}$ is meant to convey that these quantities transform like vectors under isospin rotations in the same way that $\mathbf{L}$ and $\boldsymbol{\sigma}$ transform as vectors under ordinary spatial rotations, so the rules for making isospin-invariant potentials is similar to what was found above for rotationally invariant (but spin-dependent) potentials. The main difference is that quantities like $\mathbf{L}$ could be combined with the matrices $\boldsymbol{\sigma}$ to make invariants because both transform under ordinary rotations. It is not possible to similarly combine $\vec{\tau}$ with $\mathbf{L}$ to make an invariant because although $\vec{\tau}$ is a vector under isospin rotations, it is not a vector under spatial rotations. Similarly, $\mathbf{L}$ is a rotation vector but does not transform under isospin. (This is why we use different notation to represent these two types of vectors — bold-face to denote spatial vectors but the arrow superscript to denote isovectors.) The isospin rotations are an example of an *internal symmetry* under which a label (like neutron- or proton-ness) transforms that has nothing to do with symmetries of spacetime.

We can finally specify what a charge-independent (or isospin-invariant) inter-nucleon potential looks like. Its isospin dependence can resemble (5.33), but with the vectors involved involving the only two isovectors in the problem, the isospin matrices for each of the two interacting nucleons: $\vec{T}^{(1)} := \frac{1}{2}\,\vec{\tau}^{\,(1)}$ and $\vec{T}^{(2)} := \frac{1}{2}\,\vec{\tau}^{\,(2)}$. That is, the isospin dependence of the potential must have the form

$$V(\mathbf{x}) = V_0(\mathbf{x})\, I + V_1(\mathbf{x})\, \vec{T}^{\,(1)} \cdot \vec{T}^{\,(2)}\,, \tag{5.39}$$

where we suppress any spin-dependence of these two terms. Here I is the unit matrix in the 4-by-4 isospin space spanned by the two isospin states (n and p) of the two initial and two final nucleons, and the action of $\vec{T}^{\,(1)}$ and $\vec{T}^{\,(2)}$ on two-nucleon states is given — compare with (5.30) and (5.31) — by

$$\langle f|\,\vec{T}^{\,(1)}|i\rangle = \sum_{abcd}\langle ab|\,\vec{T}^{\,(1)}_{ab;cd}\,|cd\rangle = \frac{1}{2}\sum_{abcd}\langle ab|\,\vec{\tau}_{ac}\,\delta_{bd}\,|cd\rangle\,, \tag{5.40}$$

and

$$\langle f|\,\vec{T}^{\,(2)}|i\rangle = \sum_{abcd}\langle ab|\,\vec{T}^{\,(2)}_{ab;cd}\,|cd\rangle = \frac{1}{2}\sum_{abcd}\langle ab|\,\delta_{ac}\,\vec{\tau}_{bd}\,|cd\rangle\,, \tag{5.41}$$

where now the indices a, b, c and d take values $\uparrow = p$ and $\downarrow = n$ representing the two nucleon states. These satisfy the same properties as do spin matrices when acting on a two-nucleon state: $[\vec{T}^{\,(1)}]^2 = [\vec{T}^{\,(2)}]^2 = t(t+1)\, I = \frac{3}{4}\, I$, as appropriate for isospin $t = \frac{1}{2}$. Also the total isospin, $\vec{T} = \vec{T}^{\,(1)} + \vec{T}^{\,(2)}$, squares to a matrix, $\vec{T}^2$, that has eigenvalue 2 (appropriate for $t(t+1)$ with isospin $t = 1$) when acting on the symmetric combinations: $|{\uparrow\uparrow}\rangle = |pp\rangle$, $|{\downarrow\downarrow}\rangle = |nn\rangle$ and $|{\uparrow\downarrow}\rangle + |{\downarrow\uparrow}\rangle = |pn\rangle + |np\rangle$, and takes the value 0 (as appropriate for $t = 0$) when acting on the antisymmetric combination $|{\uparrow\downarrow}\rangle - |{\downarrow\uparrow}\rangle = |pn\rangle - |np\rangle$.

The implications of the isospin-dependent term can be seen by rewriting it using the same manipulations used for spin-dependent interactions, because the isospin symmetry has

the same $SU(2)$ structure as do ordinary rotations acting on spin-half states. In particular, because

$$2\,\vec{T}^{\,(1)}\cdot\vec{T}^{\,(2)} = [\,\vec{T}\,]^2 - [\vec{T}^{\,(1)}]^2 - [\vec{T}^{\,(2)}]^2 = t(t+1) - 2\left(\frac{3}{4}\right) = \begin{cases} -3/2 & \text{if } t=0 \\ \\ +1/2 & \text{if } t=1 \end{cases}, \tag{5.42}$$

when acting on the isoscalar ($t=0$) and isovector ($t=1$) two-nucleon combinations (5.39) gives

$$V(\mathbf{x}) = \begin{cases} V_0(\mathbf{x}) - \frac{3}{4}\,V_1(\mathbf{x}) & \text{if } t=0 \\ \\ V_0(\mathbf{x}) + \frac{1}{4}\,V_1(\mathbf{x}) & \text{if } t=1 \end{cases}. \tag{5.43}$$

This shows how charge-independence can be consistent with the existence of a bound state (the deuteron) within the isosinglet antisymmetric state $|pn\rangle - |np\rangle$, without there also having to be a bound state in the isotriplet symmetric combinations, $|nn\rangle$, $|pp\rangle$ and $|pn\rangle + |np\rangle$. All that is required is for V_1 to be positive and large enough in magnitude to dominate V_0 so that V has opposite signs for these two combinations. Notice also that, when combined with the fermi statistics of the nucleon, and the symmetry of the dominant $\ell = 0$ part of the spatial wave-function, the antisymmetry of the bound isosinglet state requires the nucleon spins to be in a symmetric combination, predicting that the deuteron has spin 1 (as it does).

The appearance of

$$\vec{T}^{\,(1)}\cdot\vec{T}^{\,(2)} = \frac{1}{4}\Big[\tau_1^{(1)}\tau_1^{(2)} + \tau_2^{(1)}\tau_2^{(2)} + \tau_3^{(1)}\tau_3^{(2)}\Big], \tag{5.44}$$

implies V contains what is called an 'exchange' interaction, that interchanges n with p in addition to depending on $\mathbf{x}$. That this is present can be seen from the appearance of the matrices τ_1 and τ_2, both of which (5.38) and (5.40) and (5.41) show are off-diagonal. For instance, acting on a two-nucleon state these equations imply

$$\tau_1^{(1)}\,\tau_1^{(2)}\,|n\,p\rangle = \tau_1^{(1)}\,\tau_1^{(2)}\,|{\downarrow}{\uparrow}\rangle = |{\uparrow}{\downarrow}\rangle = |p\,n\rangle\,. \tag{5.45}$$

This is also an important feature for the internucleon potential to have, since this kind of interaction acts like a hard repulsive core that can cause strong back-scattering when nucleons interact. It is precisely this kind of interaction that is needed to account for the saturation of nuclear forces that is required for understanding nuclear properties.

Exercise 5.6: Use the explicit expressions for the matrices in Exercise 5 to write out the explicit form for the 4-by-4 matrix $\vec{T}^{(1)}\cdot\vec{T}^{(2)}$ in a two-nucleon sector with basis states given by

$$\begin{pmatrix} |p\,p\rangle \\ |n\,p\rangle \\ |p\,n\rangle \\ |n\,n\rangle \end{pmatrix}.$$

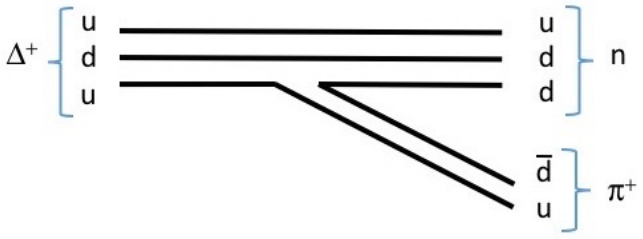


Figure 38. A quark rearrangement in which the production of a $\bar{d}d$ pair allows a Δ^+ baryon to convert to a neutron while emitting a charged pion: $\Delta^+ \to n\,\pi^+$. Similar processes also allow closely related reactions like $\Delta^- \to n\pi^-$, $\Delta^- \to p\pi^0$, $\Delta^0 \to p\pi^-$ as well as absorption processes like $p + \pi^+ \to \Delta^{++}$, $p + \pi^- \to \Delta^0$, $n + \pi^- \to \Delta^-$ and so on. Because these proceed through the strong interactions their rates are much faster than for electromagnetic and weak-interaction processes.

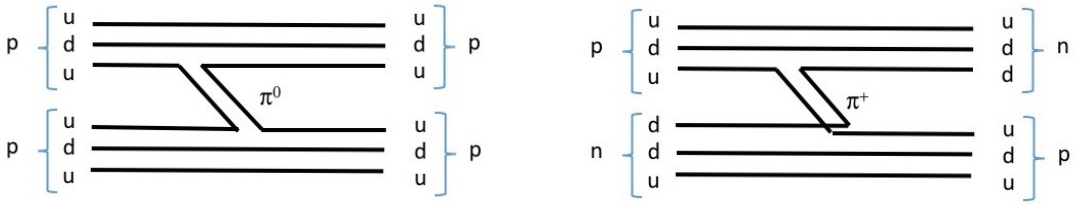
Use this to evaluate the size of the matrix elements $\langle n\,n|V|n\,n\rangle$, $\langle n\,p|V|p\,n\rangle$, $\langle n\,n|V|n\,p\rangle$ and $\langle p\,p|V|p\,p\rangle$ in terms of V_0 and V_1, for the potential given in (5.39).

5.3.3 Pions and inter-nucleon interactions

How can spin-dependent, tensor and exchange potentials arise from the underlying quark and gluon physics of the strong interactions? It happens that the longest-range part of the inter-nucleon force can be regarded as arising due to the exchange of pions between nucleons. Such an exchange is quite likely to happen for nucleons sufficiently close to one another through reactions that rearrange the quarks from which nucleons and pions are made. Figure 38 illustrates this by showing how quark-antiquark production within a baryon can cause it to emit a pion, such as the pictured reaction $\Delta^+ \to n\,\pi^+$. (The initial quark combination cannot be a proton because $p \to n\pi^+$ is not allowed by energy-momentum conservation.) The reverse reaction, wherein an antiquark annihilates one of the nucleon quarks, similarly allows a pion to be absorbed by a nucleon, such as through $\pi^+ n \to \Delta^+$. The production of Δ mesons as intermediate states makes these baryons appear as resonances when scattering pions from nucleons, such as in $\pi^- p \to \Delta^0 \to \pi^- p$. This is why they are also sometimes known as the $3-3$ resonance: they arise in pion-nucleon scattering in the spin-$\frac{3}{2}$ and (see below) isospin-$\frac{3}{2}$ channel.

While energy-momentum conservation forbids charged pions from being directly emitted by a nucleon, it does not prevent their being emitted provided they are then re-absorbed quickly (and nearby) enough to allow the uncertainty relation to interfere with energy-momentum conservation. This means that pion exchange can give rise to interactions between protons and neutrons, once these are brought sufficiently close enough together that they are separated by less than the pion's Compton wavelength, $\lambda_\pi = \hbar/m_\pi c \sim 2$ fm (which is the distance over which the uncertainty principle can cause uncertainties in energy and momenta of order m_π). The exchange of neutral pions in this way does not change the character of

the emitting (or absorbing) nucleon (as in the left panel of Figure 39), while charged pion exchange also swaps the proton with the neutron (as in the right panel of Figure 39).



(a) $pp \to pp$ through π^0 exchange

(b) $np \to pn$ through π^+ exchange

Figure 39. The quark-level processes corresponding to neutral pion exchange between nucleons (left panel) and to an exchange reaction in which $np \to pn$ through the exchange of a charged pion (right panel).

The energy change associated with several such pion emissions and absorptions occurring in sequence turns out to be responsible for the longest-range part of the inter-nucleon interaction. We see that the effective range of this part of the force is set by the pion's Compton wavelength, $R \lesssim \lambda_\pi$, and it is because the pion is the lightest meson that it dominates the long-distance part of the potential. The pion mass therefore also explains why inter-nucleon forces extend out only to distances of order a fm.

Many properties of the inter-nucleon potential can be related to properties of the pion-emission and -absorption process, which themselves can be measured directly by bombarding nucleons with pions. In particular, pion emission and absorption appears to preserve isospin,[29] with the three pion states — π^+, π^- and π^0 — transforming as an isotriplet (*i.e.* $2t+1$ states with $t = 1$) and the $2t + 1 = 4$ iso-quadruplet of Δ baryons transforming with $t = \frac{3}{2}$. In particular, as we've seen, the exchange of charged pions can generate the inter-nucleon exchange potential. Furthermore, the couplings of the isotriplet pions, $\vec{\pi}$, involve the pion momentum dotted into the nucleon spin and isospin, $(\mathbf{p}_\pi \cdot \mathbf{s})\vec{T}$, and so also can introduce a spin-dependent component to the potential.

In general inter-nucleon forces are complicated, and get more so when examined at smaller and smaller distances where (among other things) more mesons are present whose exchange can compete with pion exchange. But the picture that emerges of the inter-nucleon interaction appears broadly to have what is required to account for nuclear properties.

[29] Invariance of nuclear interactions under isospin rotations is ultimately understood in terms of the interactions between quarks and gluons that bind them into nucleons. These turn out to be invariant under isospin rotations that treat u quark as $t = \uparrow$ and d quarks as $t = \downarrow$ in the limit that these two quarks have equal masses. Since these masses are not precisely equal, small isospin-breaking effects can arise that are of order the quark masses, $m_u, m_d \sim$ few MeV, divided by the typical strong-interaction scale $\Lambda_s \sim 200$ MeV.

5.4 Radioactivity

With the previous section's general picture of nuclear binding energy in hand, we can also now better describe radioactivity in terms of nuclear decays. The general picture is that nucleons are most tightly bound in nuclei like Fe for which E_B/A is maximized. Decay options start to arise as one moves away from these special nuclei. Since the contribution of E_B to the nuclear energy (or mass) is $-E_B$, stable nuclei sit at the bottom of an energy valley (rather than at the top of an energy ridge) called the 'valley of stability'.

5.4.1 Decay overview

As one climbs the sides of the valley of stability by increasing or decreasing N/Z for fixed A, then the strong $(N-Z)^2$ dependence of the symmetry energy implies that the binding energies quickly shrink until eventually they are negative, indicating that nucleons no longer prefer to remain bound. Before this happens, even if $E_B(Z,A) > 0$ once

$$Zm_p + (A-Z)m_n - E_B(Z,A) > Zm_p + (A-Z)m_n - E_B(Z,A-1) \tag{5.46}$$

is true — and so $E_B(Z,A) < E_B(Z,A-1)$ — it can be profitable to shed just a single neutron. Similarly, once $E_B(Z,A) < E_B(Z-1,A-1)$ it can be profitable to shed a single proton. Once either of these is true the nucleus quickly sheds the excess proton or neutron to return to the valley of stability.

The locus of points in the $N-Z$ plane where nucleons can be easily shed in this way are called the *proton drip line* (on the side of excess protons) and the *neutron drip line* (on the side of excess neutrons). Because candidate nuclei beyond these drip lines can emit nucleons through the strong interaction their decays are usually very fast; with lifetimes of order strong-interaction timescales $\sim 10^{-23}$ sec. The criteria for proton and neutron drip are illustrated in Figure 40, which plots the relevant binding energies for the case $A = 12$. As this figure shows, for $A = 12$ nuclei are expected to decay very rapidly if $Z < 4$ or[30] $Z > 8$. Observations bear this out, with the encyclopaedic *Table of Isotopes* listing for $A = 12$ only ^{12}Be, ^{12}B and ^{12}N as being metastable enough to decay on millisecond timescales to ^{12}C (dominantly through β decay – as we now discuss).

Nuclei not far enough up the side of the valley to allow nucleon emission still have a way to try to lower their energy, through β-decays that can convert neutrons into protons (and vice-versa). For instance a candidate nucleus with an excess of neutrons can move to a more stable configuration through the reaction $n \to p\, e^- \overline{\nu}_e$ which increases Z by one and lowers N by one, keeping A fixed. Similarly, nucleons with an excess of protons relative to the valley of stability can lower their energy through *positron emission*: $p \to n\, e^+ \nu_e$ (sometimes called

[30] According to Fig. (40) the metastability of both $Z = 7$ and $Z = 8$ – or ^{12}N and ^{12}O – is difficult to determine because for these two cases the accuracy of eq. (5.17) is likely insufficient to decide which of $E_B(Z,A)$ or $E_B(Z-1,A-1)$ is larger.

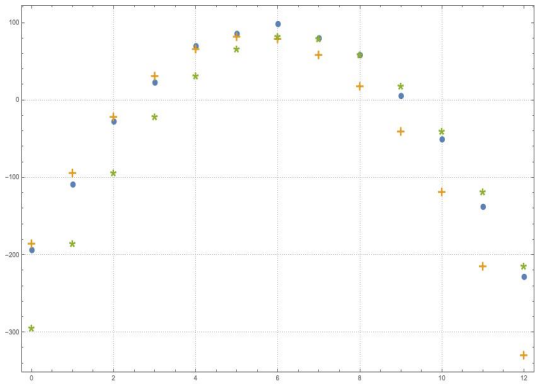


Figure 40. A plot of the binding energy $E_B(Z,A)$ (blue dots) and $E_B(Z,A-1)$ (orange + signs) and $E_B(Z-1,A-1)$ (green asterisks) vs Z for fixed $A=12$, as computed using the semi-empirical mass formula (5.17) (with numerical coefficients from (5.19)). The point where $E_B(Z,A)$ falls below $E_B(Z,A-1)$ shows where the neutron drip line intersects $A=12$ and $E_B(Z,A) < E_B(Z-1,A-1)$ does the same for the proton drip line.

β^+ decay) or by *electron capture*: $e^- p \to n\nu_e$ (both of which raise N by one and lower Z by one, at fixed A). As we shall see, all of these processes proceed only through the weak interactions and have rates that are much slower than for nucleon emission (which is why we call this new interaction 'weak').

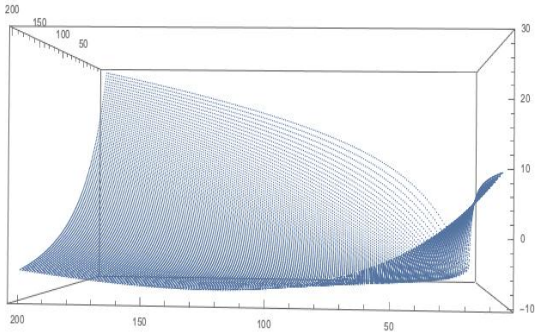


Figure 41. A 3D plot of $-E_B(Z,A)/A$ vs Z and A, as computed using the semi-empirical mass formula (5.17) (with numerical coefficients from (5.19)). The figure shows both the steep sides of the valley of stability (particularly at smaller A) and the much shallower curvature along the bottom of the valley.

Alternatively, moving along the valley of stability (rather than transverse to it) also leads to instability due to the slow falloff of E_B/A with increasing A seen in Figures 29 and 41. Once beyond Lead ($Z=82$) all nuclei become unstable, either to β-decay or to the

spontaneous emission of small ^{4}He nuclei (or α particles, which we have seen are particularly tightly bound), or (for very massive nuclei) to spontaneous *fission* into several much smaller fragments of roughly similar size. While α decays reduce both Z and N by two (and so also lower A by four), fission can reduce A and Z by much larger values in one step. But both take the nucleus back down the valley of stability towards the smaller nuclei that bind their nucleons more tightly.

Nuclear fission is the mechanism that underlies the energy release of nuclear reactors. Some nuclei are more unstable towards fission than are others, and it can be possible to initiate the process by first bombarding with neutrons, whose absorption can produce the more fissile nuclide from a more stable one. Chain reactions can then become possible because neutrons are often emitted as a by-product of the fission reaction itself. It is usually necessary to slow these emitted neutrons down because slower neutrons have much higher absorption cross sections (see *e.g.* Figure 21), and this is the reason nuclear reactors have *moderators*; relatively light materials with which energetic neutrons can scatter (and thereby be slowed down, or cooled).

Conversely, energy can also be extracted from very light nuclei, like Hydrogen or Deuterium, by fusing them together into heavier nuclei, thereby profiting from the growth of E_B/A with increasing A for small nuclei. This process — called *nuclear fusion* — is only possible if the initial nuclei can be brought together fast enough to overcome their Coulomb repulsion at larger distances so that they can get close enough together to profit from the short-ranged nuclear interactions. Collisions this fast can occur at high enough temperatures, and the fusion of Hydrogen into Helium is the source of energy that powers relatively young stars, such as our Sun.

Most often the above decays, or any nuclear reactions caused by colliding nuclei with one another or with nucleons, generate final nuclei that are not in their ground states. In this case the excited nuclei can often de-excite through the emission of a very energetic photon, or γ ray. This kind of de-excitation is responsible for γ-rays seen in radioactive materials. Because the process involved is electromagnetic (due to the need to create a photon), lifetimes for γ decays are usually longer than for intrinsically nuclear processes but shorter than for weak-interaction-mediated processes like β decays.

5.4.2 α decays

As described above, α decay is the process whereby a heavy nucleus decays (and so increases its binding energy per nucleon) through α-particle emission. We have seen that nucleons, particularly in light nuclei, tend to gather together into α-particle clusters within nuclei, so it is not so crazy that these clusters could sometimes escape and thereby fragment a nucleus. In general, energy conservation allows a nucleus to fragment with $N(Z, A) \to N(Z', A') +$

$N(Z - Z', A - A')$ once

$$E_B(Z, A) < E_B(Z', A') + E_B(Z - Z', A - A') \,. \tag{5.47}$$

Usually the very tight binding of the α particle means this criterion is satisfied first for $Z' = 2$ and $A' = 4$ and when this is true the nucleus becomes unstable towards α emission. A comparison of the energy released for several choices of (Z', A') is shown in Table 12 using the case of Uranium 232 as an example, which shows that only the α particle is so tightly bound that it can escape to infinity.

Table 12. Energy release (in MeV) if various particles were emitted from ^{232}U

Particle:	n	p	^{2}H	^{3}H	^{3}He	^{4}He	^{5}He	^{6}He	^{6}Li
Energy:	−7.26	−6.12	−10.7	−10.2	−9.92	+5.41	−2.59	−6.19	−3.79

But if the α particle can escape to infinity with positive energy, why is it trapped at all by the nuclear potential well? It remains trapped because its electric charge ensures the α particle 'sees' a Coulomb barrier when outside a nucleus, and although this Coulomb repulsion is overwhelmed by nuclear forces once within a nucleus itself, its value just outside is sufficiently large to keep α particles from leaving the nucleus (at least classically – see Figure 42). We have seen that the Coulomb energy of two protons separated by 1 fm is of order 1.4 MeV, and so the Coulomb energy of an α-particle with charge $2e$ a distance 1 fm from a decay daughter of charge $(Z - 2)e$ is $(Z - 2)2.8$ MeV, or 250 MeV for $Z = 92$ (as for Uranium). It is this large Coulomb barrier that traps a few-MeV α particle (at least classically) within a nucleus. (This large barrier is also why scattering of α particles from heavy nuclei never deviates from Rutherford scattering: the point of closest classical approach is beyond the reach of nuclear forces.)

In order to escape, an α particle must tunnel through this classical energy barrier and it is this tunnelling that is at the root of the wide diversity of lifetimes observed for α-emitters. Some of these, such as Uranium, have lifetimes in the billions of years and so can be found naturally occurring even though the nuclei are basically unstable. Other α-emitters can decay much much faster than this, by an amount that is often correlated with the energy loss released in the decay. For the even-even isotopes of any particular element the α-decay rates, Γ, are related to the energy release, Q, by the phenomenological *Geiger-Nuttall rule*:

$$\ln \Gamma(Q) \simeq -\frac{\mathcal{C}_0}{\sqrt{Q}} + \mathcal{C}_1 \,, \tag{5.48}$$

where $\mathcal{C}_0$ and $\mathcal{C}_1$ are constants that differ for different chemical elements but are the same for the isotopes of any specific element. This relation is plotted in Figure 43, together with

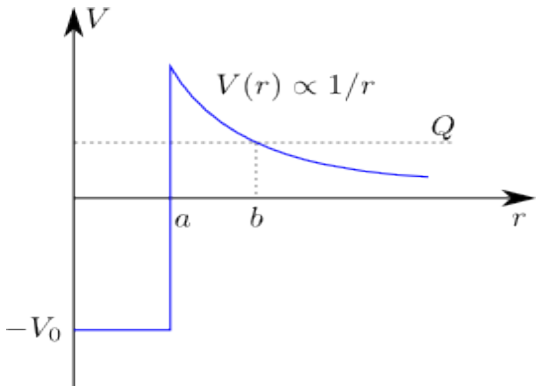


Figure 42. A sketch of the potential barrier through which an α-particle tunnels. (Figure source: `http://physics.stackexchange.com/questions/102364/what-does-the-coulomb-barrier-really-mean`)

measured values for experimental lifetimes and energy loss. The figure shows how the exponential dependence of Γ on Q is such that a factor of 2 in Q can correspond to many orders of magnitude in decay width.

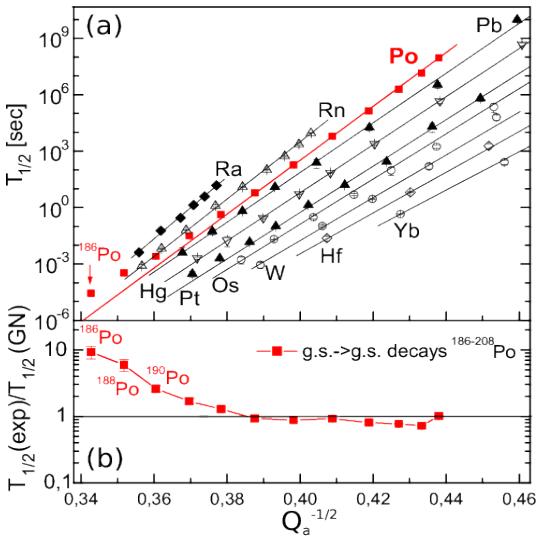


Figure 43. A plot of the α-decay width vs the energy loss Q for several isotopes of even-even nuclei. (Figure source: `Qi et. al., Physics Letters B734 (2014) 203 (arXiv:1405.5633).`)

Tunnelling can account for such a wide range of lifetimes because tunnelling rates depend exponentially on the shape of the potential being tunnelled under, with rate $\Gamma \simeq (v_{\text{in}}/R)e^{-2G}$ with $v_{\text{in}} = \sqrt{2M(Q+V_0)}$ being the speed of the α-particle within the nuclear well (modelled as a square well with potential $-V_0$ within the range $r < R$) and so v_{in}/R gives the rate with which it classically arrives at the inner side of the barrier. The exponential factor comes from

the tunnelling probability, with the *Gamow factor*, G, given (for zero angular momentum, ℓ) by

$$G = \int_a^b \mathrm{d}r \, \sqrt{2M[V(r) - Q]} \,. \tag{5.49}$$

Here M is the alpha-particle mass, while $V(r) = 2(Z-2)\alpha/r$ is the Coulomb potential for an α-particle and the daughter nucleus (with charge $Z-2$). The integration limits are: $a = R$, the radius of the nuclear potential well, and $b = 2(Z-2)\alpha/Q$, the classical closest approach for an α-particle of energy Q (see Figure 42).

Evaluating the integrals gives

$$\begin{aligned} G &= 2\alpha(Z-2)\sqrt{\frac{2M}{Q}}\left[\cos^{-1}\sqrt{x} - \sqrt{x(1-x)}\right] \\ &\simeq 2\alpha(Z-2)\sqrt{\frac{2M}{Q}}\left[\frac{\pi}{2} - 2\sqrt{x} + \cdots\right] \quad (\text{when } x = Q/V_C(R) \ll 1)\,, \end{aligned} \tag{5.50}$$

with $x = a/b = Q/V_C(a)$, where $V_C(a) = 2(Z-2)\alpha/a$ is the height of the Coulomb barrier at $r = a = R$. The approximate equality in the expression for G gives the leading behaviour when $x \ll 1$ (which is the limit of physical interest). This clearly reproduces the exponential dependence on Q appearing in the Geiger-Nuttall relation, with

$$\mathcal{C}_0 = \pi\alpha(Z-2)\sqrt{2M} \qquad \text{and} \qquad \mathcal{C}_1 \simeq 4\sqrt{\alpha(Z-2)MR}\,. \tag{5.51}$$

In particular notice $(\mathcal{C}_1/\sqrt{Q})/\mathcal{C}_0 \simeq (\pi/4)\sqrt{2\alpha(Z-2)/QR} \gg 1$ because $Q \ll V_C(a) = 2\alpha(Z-2)/R$, so that relatively small changes in Q cause enormous changes to Γ. We see in this way how the interplay between the strong and electromagnetic interactions can account for both the dramatic range of possible α-decay lifetimes and (in some circumstances) their correlation with the emitted α-particle energy.

5.4.3 β decays and multiple neutrinos

Whereas α and γ decays are well-described in terms of nucleons interacting through the electromagnetic and strong interactions, an understanding of β decays required both the proposal of a new type of particle (the neutrinos) and the discovery of new interactions: the weak interactions. An entirely new type of interaction turned out to be required because the strong and electromagnetic interactions both preserve the number of each type of quark (and lepton). As a result they cannot describe a process whereby d and u quarks (or neutrons and protons) inter-convert into one another.

As mentioned earlier, β decays presented a puzzle to early researchers because of the continuous distribution of produced electrons. In those days it was known that a neutron converts to a proton plus an electron (as required by electric-charge conservation), through a reaction that was assumed to have the form $n \to p\, e^-$. But because this is a two-body decay

conservation of 4-momentum then implies the proton and electron should have a unique energy in the decaying neutron's rest frame, and this is not what is observed. The outgoing electron is instead seen to take a continuous range of energies in the decay, and only the uppermost energy of this range equals the difference of binding energies between the initial and final nucleus.

This puzzle led to Pauli's proposal of the existence of a new particle that also would appear in these decays. To have avoided discovery it would have to be electrically neutral (and, like the electron, not participate in nuclear interactions) and to appear in the decay it would have to be much lighter than the neutron; hence[31] the name neutrino.

Over time, with the discovery of more particles and their decays, two more things became clear. The first was that the same kind of weak interaction needed to describe nuclear β-decay could also account for the decay of these new particles. The second was that more than one neutrino might be necessary. The crucial clue for the need for more neutrinos came from the *absence*[32] of interactions that should have been present were only one neutrino involved.

To see why more neutrinos were required, it is important to recognize that although neutrinos themselves are close to invisible (because they interact so weakly), they can be produced in beams intense enough to measure their presence. And when they are produced they are always produced in association with a charged lepton. For instance, neutrino beams are in practice produced by reactions like $n \to p + e^- + (\text{invisible})$ and the 'invisible' flux can be large enough for the neutrinos in it to be detected downstream if the neutron decays are produced within the intensely radioactive core of a nuclear reactor. Alternatively, neutrino beams can be produced from the reaction $\pi^\pm \to \mu^\pm + (\text{invisible})$ and this has a large flux if the pions are caused by strong nuclear reactions (such as by bombarding nuclei with protons, neutrons or α-particles) because these emit pions and every pion decays essentially 100% of the time to muons plus neutrinos. Essentially every muon produced this way then also decays to an electron plus more neutrinos: $\mu^\pm \to e^\pm + (\text{invisible})$.

Most importantly, the reactions used to detect the presence of the neutrinos downstream of such a production process also usually involve a charged lepton, such as if they are detected through the inverse processes, like $\nu + n \to p + e^-$ or $\nu + n \to p + \mu^-$. What (until very recently) was *never* seen to happen was to have a neutrino produced in association with an electron (in a reactor, say) then be detected in association with a muon (or vice versa). This led to the hypothesis of there being two types of neutrinos, ν_e and ν_μ, whose interactions would separately conserve electron and muon number: L_e and L_μ. Only ν_e takes part in reactions associated with electrons and only ν_μ takes part in interactions associated with muons. Both must be present in the reaction $\mu^\pm \to e^\pm + (\text{invisible})$, and the presence of two invisible particles is in this case also indicated by the fact that the outgoing electron (or

[31] The ending '-ino' being deminuative in Italian.

[32] Much like Sherlock Holmes' ability to solve a crime using the clue of the dog that *didn't* bark in the night.

positron) has a continuous distribution of energies rather than the unique energy that would have been required by a two-body decay.

5.4.4 Neutrino oscillations

This picture of multiple neutrinos and the separate conservation of L_e and L_μ provided a very good description of all neutrino measurements for more than 50 years, until it began to unravel not so long ago with the discovery of neutrino oscillations. In retrospect, the first signs of a problem arose when people began to measure the neutrino flux coming to us from the Sun. These neutrinos are produced by the same chain of nuclear reactions that power the Sun, whose net result is

$$2\,e^- + 4\,p \to {}^4\mathrm{He} + 2\,\nu_e\,, \tag{5.52}$$

which releases energy because the Helium nucleus is so tightly bound. The two electrons are required by electric charge conservation and so the two neutrinos are required by lepton number conservation. They must be electron neutrinos because the underlying reaction at work is $p + e^- \to n + \nu_e$, and so occurs in association with an electron.

Since we know how much energy comes out of the Sun we know how many reactions took place and so can work out the number of neutrinos being sent our way. The problem is that once these neutrinos were detected there were never quite as many as expected. Because the detectors dominantly detected the neutrinos using interactions associated with electrons, such as quasi-elastic scattering from protons: $\nu_e\,p \to n\,e^+$, they were mostly just sensitive to the presence of the one type of neutrino. This was fine so long as only the one type of neutrino comes from the Sun, but the persistent shortfall eventually raised strong doubts about whether this was really so. The issue was eventually settled with the development of a detector — the *Sudbury Neutrino Observatory*, or SNO — that could detect all three species of neutrino. SNO verified that the expected neutrino flux really is there, it just involves all three of the known neutrino species and not just ν_e.

Around the same time further evidence that L_e and L_μ are not separately conserved also became available from the study of neutrinos produced when cosmic rays hit the Earth's upper atmosphere. Cosmic rays are mostly protons and when these hit Nitrogen and Oxygen nuclei in the atmosphere they emit many pions. Because essentially every charged pion decays through the process

$$\pi^+ \to \mu^+ + \nu_\mu \to (e^+ + \nu_e + \overline{\nu}_\mu) + \nu_\mu\,, \tag{5.53}$$

or its counterpart with $\pi^- \to \mu^- \to e^-$, it was expected that one would find close to two muon neutrinos (or anti-neutrinos) for every electron-type neutrino. But although experiments found this to be largely true for neutrinos coming down from collisions in the atmosphere overhead, they instead found that when the neutrinos are detected coming up (starting from a collision in the atmosphere on the opposite side of the Earth, with the neutrinos penetrating through the entire Earth — which is possible because neutrinos interact so weakly) the ratio

is more like one-to-one. Again the total number of neutrinos seemed OK but the distribution into each species was not consistent with conservation of L_e and L_μ.

Both these lines of discovery have since been also verified using man-made (and so better controlled) neutrino beams, rather than extra-terrestrial sources. Although separate conservation of L_e and L_μ can be a very good approximation in many circumstances, we now have good experimental evidence for two things: (*i*) there are indeed (at least) three species of neutrino,[33] and (*ii*) it is clear that L_e and L_μ are not exactly conserved.

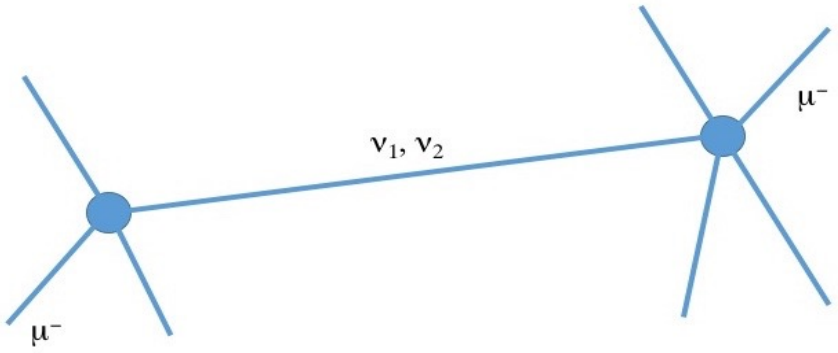


Figure 44. A process where a charged muon interacts with something to produce a neutrino, which then propagates to react with other particles and produce another charged muon.

It turns out that there is a nonzero amplitude to find all species of neutrinos emitted in association with any particular charged lepton. This turns out only to be observable when neutrinos propagate over relatively long distances because then a pattern of interference ('oscillations') can be seen in the admixture of different neutrinos. For instance, suppose an initial neutrino-production event involving an associated charged muon takes place at time t_0 with an amplitude, a_i, with $i = 1, 2$ corresponding to each of two neutrino types ν_1 and ν_2 (the generalization to more neutrinos is straightforward). We do not label the neutrino species with e or μ anymore because in the general case the neutrinos are not associated exclusively with charged electrons or muons. The amplitude for the neutrino to be absorbed at some later time t in association with another muon, say, is similarly a_i^*. For only two types of neutrinos a_i satisfies $\sum_i |a_i|^2 = |a_1|^2 + |a_2|^2 = 1$, since either one type or the other type

[33] Part of this evidence comes from measuring the rate for Z bosons to decay 'invisibly', which can be determined by measuring their total decay width. (This can be done despite some of the decays being invisible by measuring the total width, Γ, of the Z-boson resonance shown in Fig. 11, for example.) Comparing the result with the rate for observed Z decays shows that the Z indeed sometimes does decay invisibly, with a rate consistent with $Z \to \nu\overline{\nu}$ provided there are 3 (not 2 or 4) neutrino species participating in the decay, whose masses are much lighter than half the Z mass. More than 3 neutrinos can exist only if they either do not couple to the Z boson, or are too heavy to appear in these decays.

must be emitted with total probability 1. As a result we can write $a_1 = \sin\theta$ and $a_2 = \cos\theta$, for some angle, θ, called the *neutrino mixing angle*.

The time-dependence of the joint amplitude for neutrino production followed by absorption then is

$$\mathcal{A}_{\mu\to\mu}(t,t_0) = \mathcal{A}_0 \sum_i |a_i|^2\, e^{-iE_i(t-t_0)} = \mathcal{A}_0 \Big[\sin^2\theta\, e^{-iE_1(t-t_0)} + \cos^2\theta\, e^{-iE_2(t-t_0)}\Big], \qquad (5.54)$$

where $\mathcal{A}_0$ is the amplitude for the rest of the process and the phases are due to the evolution of the neutrino state from t_0 to t, where $E_i = \sqrt{p^2 + m_{\nu_i}^2}$ and m_{ν_i} is the mass of neutrino type 'i'. In these expressions the factors of a_i come from the amplitude for emitting the neutrino and a_i^* comes from the amplitude for this neutrino's later absorption. The total probability for this process therefore is

$$\begin{aligned} P_{\mu\to\mu}(t,t_0) = |\mathcal{A}(t,t_0)|^2 &= |\mathcal{A}_0|^2 \Big\{\cos^4\theta + \sin^4\theta + 2\sin^2\theta\cos^2\theta\cos\Big[(E_2-E_1)(t-t_0)\Big]\Big\} \\ &= |\mathcal{A}_0|^2 \Big\{1 - 2\sin^2\theta\cos^2\theta + 2\sin^2\theta\cos^2\theta\cos\Big[(E_2-E_1)(t-t_0)\Big]\Big\} \\ &= |\mathcal{A}_0|^2 \left[1 - \sin^2(2\theta)\sin^2\left(\frac{\Delta m^2 L}{4E}\right)\right], \end{aligned} \qquad (5.55)$$

where we use $E_i \simeq p + m_{\nu_i}^2/2p + \cdots \gg m_{\nu_i}$ and $t - t_0 \simeq L/c = L$ for ultra-relativistic neutrinos (where L is the distance between the neutrino-production and absorption events) to write $\frac{1}{2}(E_2 - E_1)(t - t_0) \simeq (m_{\nu_2}^2 - m_{\nu_1}^2)(t - t_0)/(4p) \simeq \Delta m^2 L/(4E)$. Here Δm^2 denotes the difference between the neutrino squared-masses. The probability for starting with muon-associated neutrino production event and absorbing the neutrino in an electron-associated event is similarly

$$P_{\mu\to e}(t,t_0) = |\mathcal{A}_0|^2 \left[\sin^2(2\theta)\sin^2\left(\frac{\Delta m^2 L}{4E}\right)\right], \qquad (5.56)$$

so that their sum is $|\mathcal{A}_0|^2$, as it must be.

The name 'oscillations' comes from the oscillatory dependence of these expressions on the production-to-detection distance, L. The distance $\lambda := 4E/\Delta m^2$ is called the oscillation length, since it shows how big L must be in order to see an appreciable effect. For $\Delta m^2 \simeq 10^{-3}$ eV^2 and $E \sim 1$ GeV the oscillation length is $\lambda \sim 10^{12}$ eV$^{-1} \sim 100$ km, showing that production and detection must take place significantly far apart from one another to see oscillations using neutrinos produced on Earth.

> **Exercise 5.7:** For two neutrino species repeat the arguments made when establishing eqs. (5.54) and (5.55) but for $\mu \to e$ processes (for which the neutrino is produced in association with a muon and detected in association with an electron), and thereby provide an independent derivation of (5.56).

5.4.5 β decays and the weak interactions

Weak decays are usually much slower than decays that proceed through the electromagnetic and strong interactions. Their rate is characterized by a new interaction constant of nature, called the Fermi constant: $G_F = 1.166379 \times 10^{-5}$ GeV^{-2}. This was first captured in the *Fermi* theory of β-decay, with further refinements being added with the development of the '$V-A$' theory of the weak interactions by Feynman and Marshak and by Sudarshan. These theories have since been superseded by the Standard Model of particle physics (as described below in more detail), within which the earlier theories arise instead as the low-energy limit of a more fundamental interaction involving the exchange of a new particle: the W boson.

For instance for muon decay, $\mu^-(k) \to e^-(p)\overline{\nu}_e(q)\nu_\mu(r)$, the invariant amplitude is predicted in these theories to be

$$\mathcal{M} = 64 G_F^2 (p\cdot r)(k\cdot q)\,, \tag{5.57}$$

where $p\cdot r = \eta_{\mu\nu}p^\mu r^\nu$ and so on. Using this in (2.29) then gives the differential decay rate. Typically only the outgoing electron momentum is observed in this decay, allowing one to perform the sum and integral over the other (unmeasured) final-state spins and momenta. For unpolarized muons the electron direction is isotropic and after performing the integrals over unmeasured momenta the differential rest-frame rate as a function of the electron energy becomes

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}\varepsilon} = \frac{G_F^2 m_\mu^5}{4\pi^3}\left(\varepsilon - \frac{4\varepsilon^2}{3} + \varepsilon\, r^2 - \frac{2r^2}{3}\right)\sqrt{\varepsilon^2 - r^2}\,, \tag{5.58}$$

where $\varepsilon := E_e/m_\mu$ and $r := m_e/m_\mu$. This is a monotonically increasing function and so the most likely electron energy is the maximum available: $\varepsilon_{\max} = \frac{1}{2}(1+r^2)$ (corresponding to one of the neutrinos carrying away no energy at all).

Performing the last integration over the range $r \le \varepsilon \le \frac{1}{2}(1+r^2)$, the total rest-frame muon decay rate then finally becomes

$$\Gamma(\mu \to e\overline{\nu}\nu) \simeq \frac{G_F^2 m_\mu^5}{192\pi^3}\, J(r)\,, \tag{5.59}$$

where the mass-dependent function, $J(r)$, is given by

$$J(r) = 1 - 8r^2 - 24r^4 \ln r + 8r^6 - r^8\,, \tag{5.60}$$

and differs from unity at the 10^{-4} level given the mass ratio $r = m_e/m_\mu \simeq 0.511/105 \sim 5\times 10^{-3}$. It is comparisons of this expression (plus some sub-leading corrections) with precise measurements of the muon lifetime that give our best value for G_F.

> **Exercise 5.8:** Use (5.57) in (2.29) to derive (5.58). Sketch your result for $\mathrm{d}\Gamma/\mathrm{d}\varepsilon$ as a function of ε. Integrate (5.58) to derive (5.59) and (5.60).

Notice that the mean lifetime predicted by this formula is $\tau_{\rm wk} = 1/\Gamma$ of order microseconds, which is much longer than the typical strong-interaction decay time, $\tau_{\rm str} \sim 1/m_\mu$ which is of order 10^{-23} seconds. This huge difference is largely driven by the small dimensionless ratio $G_{\scriptscriptstyle F} m_\mu^2 \sim 10^{-7}$. This turns out to be a general conclusion: weak interactions are weak provided the energy release involved remains small compared with $1/\sqrt{G_{\scriptscriptstyle F}} \sim 300$ GeV. At energies much larger than this the nature of the weak interactions turns out to change, in a way that more strongly resembles electromagnetic interactions in their strength. The story of this change, leading to a unified picture of *electro-weak* interactions at these higher energies, is told by the Standard Model of particle physics.

As is now illustrated with a few examples, the same coupling, $G_{\scriptscriptstyle F}$, also gives a good description of other decays, including nuclear β-decay rates. (One of these has already been described, in (2.34).) For *hadronic* systems involving quarks there are two complications to this comparison. First, there is a small correction because quarks do not quite enjoy precisely the same weak-interaction couplings as do leptons (more about which later). But another complication when considering hadronic (and in particular nuclear) decays is the uncertainty of their structure, since this is poorly understood and often cannot be computed reliably. An important exception where the nuclear structure does *not* interfere with β-decay calculations arises for decays between two spin-zero states that lie within the same isospin multiplet, because in this case it turns out that the invariant amplitude is dictated by isospin symmetry considerations. Concrete examples of such decays are the super-allowed decay $^{14}{\rm C} \to {}^{14}{\rm N}$ (relevant to radio-carbon dating), where the decay is to the isospin partner of the ^{14}C ground state. Another process of this type is the pion decay $\pi^- \to \pi^0\, e^-\, \overline{\nu}_e$.

Exercise 5.9: Consider the case of super-allowed β-decay of the form $A(p) \to B(q) + e^-(k) + \overline{\nu}_e(r)$, where p^μ, q^μ, k^μ and r^μ are the respective 4-momenta and the initial and final hadron (nucleus, baryon or meson) are partners within a single isospin multiplet. This includes the two examples $^{14}{\rm C} \to {}^{14}{\rm N}$ and $\pi^- \to \pi^0\, e^-\, \overline{\nu}_e$ mentioned in the main text. In this case the invariant matrix element for the decay $A \to B + e^- + \overline{\nu}_e$ turns out to be

$$16 G_{\scriptscriptstyle F}^2 C^2 |V_{ud}|^2 \Big[2(k\cdot q)(r\cdot q) - q^2 (k\cdot r) \Big] \,, \tag{5.61}$$

where the Kobayashi-Maskawa matrix element, $|V_{ud}| = 0.974$, quantifies the relative strength of the weak interactions for $u-d$ quarks relative to leptons (and is determined by comparing our final prediction for Γ with experimental measurements). The quantity C depends on the isospin quantum numbers of the states A and B by

$$C = \sqrt{(t - t_3)(t + t_3 + 1)} \,, \tag{5.62}$$

where t is the isospin of the common isospin multiplet in which both A and B live, and $t_3 = t_3(A) = t_3(B) - 1$ is the eigenvalue of the 3rd component of isospin

for these states. For example, for both ^{14}C and π^- decay the parent and daughter lie in an isotriplet multiplet, with $t = 1$ and $t_3 = t_3(A) = -1$ and $t_3(B) = 0$ so $C = \sqrt{2}$.

After integrating over the unmeasured neutrino momentum and the momentum of the recoiling daughter particle B, show that the rest-frame differential decay rate as a function of electron momentum is given by

$$\mathrm{d}\Gamma[A(p) \to B(q) + e^-(k) + \overline{\nu}_e(r)] = \frac{G_F^2 C^2 |V_{ud}|^2}{2(2\pi)^4} \left\{ \frac{[-(k-p)^2 - m_B^2]^2}{-(k-p)^2} \right\} \mathrm{d}^3\mathbf{k}\,, \tag{5.63}$$

where $-(k-p)^2 = m_A^2 + m_e^2 - 2E_e m_A$ in the decay rest-frame. This shows the rate is independent of the electron direction (after integrating over the neutrino and recoil directions). Finally, defining $Q = m_A - m_B$ as the energy released by the decay, and neglecting both m_e/m_A and Q/m_A (but not necessarily m_e or E_e relative to Q), show that the rest-frame differential decay rate becomes

$$\begin{aligned} \frac{d\Gamma}{dE_e} &\simeq \frac{G_F^2 C^2 |V_{ud}|^2}{(2\pi)^3} \left[\frac{(2E_e m_A + m_B^2 - m_A^2)^2}{m_A^2 - 2E_e m_A} \right] E_e \sqrt{E_e^2 - m_e^2} \\ &\simeq \frac{G_F^2 C^2 |V_{ud}|^2}{2\pi^3} \left(\frac{m_A E_e}{m_A - 2E_e} \right) (E_e - Q)^2 \sqrt{E_e^2 - m_e^2}\,. \end{aligned} \tag{5.64}$$

Neglecting the electron mass show that the allowed range for E_e is $m_e \simeq 0 \le E_e \le E_{\max} := (m_A^2 - m_B^2)/(2m_A) \simeq Q$. Perform the final integration and show

$$\Gamma(A \to B e^- \overline{\nu}_e) = \frac{G_F^2 C^2 |V_{ud}|^2 m_A^5}{2\pi^3}\, I(\Delta)\,, \tag{5.65}$$

where $\Delta := E_{\max}/m_A = (m_A^2 - m_B^2)/(2m_A^2) \simeq Q/m_A$ and

$$\begin{aligned} I(\Delta) &= -\frac{1}{4}\left[\frac{\Delta}{4} - \frac{3\Delta^2}{4} + \frac{\Delta^3}{3} + \frac{\Delta^4}{6} + \frac{1}{2}\left(\Delta - \frac{1}{2}\right) \ln\Big(1 - 2\Delta\Big) \right] \\ &\simeq \frac{\Delta^5}{30}\left[1 - \frac{3\Delta}{2} + \mathcal{O}(\Delta^2) \right]. \end{aligned} \tag{5.66}$$

Comparison of (5.65) (for a variety of super-allowed nuclear β-decays) with observations is what determines $|V_{ud}|$.

Exercise 5.10: Expression (5.65) found in Exercise 5.9 also describes well $\pi^- \to \pi^0 e^- \overline{\nu}$ decay, which (with $C^2 = 2$ for pion decay) implies

$$\Gamma(\pi^- \to \pi^0 e^- \overline{\nu}_e) = \frac{G_F^2 |V_{ud}|^2 Q^5}{30\pi^3} \left(1 - \frac{3Q}{2m_{\pi^-}} + \cdots \right). \tag{5.67}$$

As seen in (2.34) — since $\Gamma(\pi^- \to \mu^- \overline{\nu}) = \Gamma(\pi^+ \to \mu^+ \nu)$ — the same combination $G_{\scriptscriptstyle F}|V_{ud}|$ also controls the rate of $\pi^- \to \mu^- \nu$ decay. In particular, the ratio of (5.67) to (2.34) (called the *branching ratio*) gives the fraction of times a charge pion decays to leptons rather than semi-leptonically. Use this to show

$$B = \frac{\Gamma(\pi^- \to \pi^0 e^- \overline{\nu})}{\Gamma(\pi^- \to \mu^- \nu)} \simeq \frac{2Q^5}{15\pi^2 m_\mu^2 F_\pi^2 m_{\pi^-}} \left(1 - \frac{m_\mu^2}{m_{\pi^-}^2}\right)^{-2} \simeq 1.09 \times 10^{-8}\,, \quad (5.68)$$

which agrees well with the experimental value: $B_{\rm exp} = (1.036 \pm 0.006) \times 10^{-8}$ (with the difference captured by the dominant subleading theoretical corrections). The small size of $G_{\scriptscriptstyle F} Q^2$ clearly gets the size of this decay right as well.

Exercise 5.11: Neutron decay is the decay of the simplest 'nucleus.' Theory predicts the invariant differential rate for neutron decay — with 4-momentum assignments $n(r) \to p(p) + e(k) + \overline{\nu}(q)$ — is given to good approximation (in the neutron rest frame) by

$$\mathcal{M}(n \to p\, e \overline{\nu}) \simeq 16 G_{\scriptscriptstyle F}^2 |V_{ud}|^2 m_p m_n E_\nu E_e \left[\mathcal{F}\left(1 + \frac{\mathbf{q}\cdot\mathbf{k}}{E_\nu E_e}\right) + \mathcal{G}\left(1 - \frac{\mathbf{q}\cdot\mathbf{k}}{3\, E_\nu E_e}\right)\right], \quad (5.69)$$

where we neglect Q/m_p where $Q = m_n - m_p$ is the energy release, and this allows us to drop the kinetic energy of the daughter proton, so $E_p \simeq m_p + \mathbf{p}^2/2m_p + \cdots \simeq m_p$. The constants $\mathcal{F} = g_{\scriptscriptstyle V}^2$ and $\mathcal{G} = 3\, g_{\scriptscriptstyle A}^2$ are the 'Fermi' and 'Gamow-Teller' terms, arising from the $q^2 \to 0$ limit of the form factors relevant to the weak interactions (similar in spirit to the form factors used earlier for the proton's electromagnetic interactions). Both are associated with the unknown details of the substructure of the nucleon. It happens that the electric charge of the proton determines $g_{\scriptscriptstyle V} = 1$ (much as was true for $G_{\scriptscriptstyle E}(q^2 = 0)$) but $g_{\scriptscriptstyle A}$ need not also be close to unity (much as also was $G_{\scriptscriptstyle M}(q^2 = 0)$). The value of $g_{\scriptscriptstyle A}$ can be determined by comparing with the measured neutron lifetime, and because $g_{\scriptscriptstyle A}$ also appears in other expressions (like the amplitude for pion emission by nucleons) the theory is predictive.

Show that the rest-frame differential rate for $n \to p e^- \overline{\nu}$ is therefore given by

$$\frac{\mathrm{d}\Gamma(p \to p e \overline{\nu})}{\mathrm{d}^3\mathbf{p}\, \mathrm{d}^3\mathbf{q}\, \mathrm{d}^3\mathbf{k}} = \frac{G_{\scriptscriptstyle F}^2 |V_{ud}|^2}{2(2\pi)^5} \left[\mathcal{F}(1 + \mathbf{q}\cdot\mathbf{k}) + \mathcal{G}\left(1 - \frac{\mathbf{q}\cdot\mathbf{k}}{3}\right)\right] \delta^3(\mathbf{p}+\mathbf{q}+\mathbf{k})\delta(Q - E_e - E_\nu)\,. \quad (5.70)$$

Since nothing in the integrand depends on $\mathbf{p}$ its integral can be done using the momentum-conserving delta function, leaving the integrals over $\mathbf{k}$ and $\mathbf{q}$ uncorrelated in direction. Use this to perform the integration over $\mathbf{q}$ and the direction of

k to obtain the differential decay rate as a function only of electron energy:

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E_e}(n \to pe\overline{\nu}) = \frac{G_F^2 |V_{ud}|^2 (\mathcal{F}+\mathcal{G})}{2\pi^3}\, E_e(Q-E_e)\sqrt{E_e^2 - m_e^2}\sqrt{(Q-E_e)^2 - m_\nu^2}\,, \tag{5.71}$$

where $\mathcal{F}+\mathcal{G} = g_V^2 + 3g_A^2$ and m_ν is the mass of the relevant neutrino. The presence of a neutrino mass can be tested by plotting $y = \left[(\mathrm{d}\Gamma/\mathrm{d}E_e)/E_e\sqrt{E_e^2-m_e^2}\right]^{1/2}$ against E_e (called a *Kurie plot*), since this ought to be a straight line if $m_\nu = 0$.

Dropping the neutrino mass (which is at most of order an eV or so), perform the integral over $m_e \le E_e \le Q$ to obtain

$$\Gamma(n \to pe\overline{\nu}) = \frac{G_F^2 |V_{ud}|^2 (g_V^2 + 3g_A^2) Q^5}{60\pi^3}\, I\left(\frac{Q}{m_e}\right)\,, \tag{5.72}$$

with

$$\begin{aligned} I(Q/m) &= \left(1 - \frac{9m^2}{2Q^2} - \frac{4m^4}{Q^4}\right)\sqrt{1-\frac{m^2}{Q^2}} + \frac{15m^4}{2Q^4}\ln\left(\frac{Q+\sqrt{Q^2-m^2}}{m}\right) \\ &\simeq 1 + \mathcal{O}\left(\frac{m^2}{Q^2}\right) \qquad (\text{if } m \ll Q)\,. \end{aligned} \tag{5.73}$$

Again it is the quantity $G_F Q^2 \sim 10^{-11}$ that determines the decay lifetime and makes it so long. Use $g_V = 1$ and $g_A \simeq 1.267$ to evaluate the free-neutron lifetime. (You should find around 950 seconds, which is longer than the measured value of 880.3 ± 1.1 seconds. This difference is mostly to do with the distortion of the outgoing electron wave-function at $r = 0$ by the Coulomb field of the final proton, which is energy-dependent and so should be included before performing the integration over E_e.)

6 Quantum Field Theory

An important missing step in the above story is the way one computes the invariant rates, $\mathcal{M}$, that govern the cross sections and decay rates described above. Although we know how to compute this in terms of an underlying Hamiltonian for scattering in single-particle nonrelativistic quantum mechanics we do not yet know how to do this for something like a decay process, which changes the number and type of particles present.

Filling in this step (at least partially) is the goal of this section, and involves describing the formalism of quantum field theory (which is the natural language for describing processes that involve many particles (and in particular can change the number or type of particles).

6.1 Heisenberg's harmonic oscillator

Before starting, first a brief but useful aside to review the Heisenberg treatment of the one-dimensional harmonic oscillator. This is useful to review because the harmonic oscillator shares the spectrum of the quantum field theory of non-interacting many-particle systems. They resemble one another because both systems involve energy levels that are precisely equally spaced: $E_{n+1} - E_n = \omega$ is independent of n.

The single-particle 1D harmonic oscillator is defined by the time-independent Schrödinger equation

$$H\psi_n(x) = \left[-\frac{1}{2m}\frac{\partial^2}{\partial x^2} + \frac{1}{2}\,m\,\omega^2\,x^2\right]\psi_n(x) = E_n\;\psi_n(x)\,, \tag{6.1}$$

where m is the particle mass and ω is the oscillator frequency. This has eigenvalues

$$E_n = \left(n + \frac{1}{2}\right)\omega\,, \tag{6.2}$$

with eigenfunctions

$$\psi_n(x) = \sqrt{\frac{m\,\omega}{\pi 2^n n!}}\;\mathcal{H}_n(x)\,e^{-\frac{1}{2}\,m\omega x^2}\,, \tag{6.3}$$

where $n = 0, 1, 2, \ldots$ and $\mathcal{H}_n(x)$ are the (nth order) Hermite polynomials.

Heisenberg's treatment of this problem focusses on the *ladder* operator

$$A := \frac{1}{\sqrt{2m\omega}}\;(m\omega\,X + i\,P) = \frac{1}{\sqrt{2m\omega}}\;\left(m\omega\,x + \frac{\partial}{\partial x}\right)\,, \tag{6.4}$$

and its adjoint,

$$A^\star := \frac{1}{\sqrt{2m\omega}}\;(m\omega\,X - i\,P) = \frac{1}{\sqrt{2m\omega}}\;\left(m\omega\,x - \frac{\partial}{\partial x}\right)\,, \tag{6.5}$$

rather than the position and momentum operators X and P. The commutation relations $[X, P] = i$ implies A satisfies the commutation relation

$$\Big[A, A^\star\Big] = A\,A^\star - A^\star A = 1\,, \tag{6.6}$$

as can be checked by acting on an arbitrary function, $\psi(x)$, and performing the differentiations explicitly.

Since X and P can be rewritten in terms of A and $A^\star$,

$$X = \frac{1}{\sqrt{2m\omega}}\Big(A^\star + A\Big) \quad \text{and} \quad P = i\sqrt{\frac{m\omega}{2}}\Big(A^\star - A\Big)\,, \tag{6.7}$$

the same is true of any other observable for the harmonic oscillator, making A, $A^\star$ an equivalent basis of operators to X and P when describing harmonic oscillator observables. In particular, the Hamiltonian itself is given by

$$H = \frac{P^2}{2m} + \frac{1}{2}\,m\,\omega^2\,X^2 = \frac{\omega}{2}\,(A^\star A + A\,A^\star) = \omega\left(A^\star A + \frac{1}{2}\right)\,, \tag{6.8}$$

and the last equality uses the commutation relation (6.6). This expression for H is also easily verified by directly writing out the right-hand side in terms of the explicit derivatives given in (6.4) and (6.5).

The above formula for H shows in particular that the combination $N := A^\star A$ is diagonal in the energy eigenbasis with eigenvalue n. That is, if $\psi_n(x) = \langle x|n\rangle$ for the energy eigenstates $|n\rangle$, then

$$A^\star A\,|n\rangle = n\,|n\rangle \quad \text{so that} \quad H\,|n\rangle = \left(n + \frac{1}{2}\right)\omega\,|n\rangle\,. \tag{6.9}$$

This last equation implies in particular that $A^\star A$ gives zero when acting on the ground state, and this is easily verified by applying A directly to $\psi_0(x)$:

$$\langle x|A|0\rangle = \frac{1}{\sqrt{2m\omega}}\left(m\omega\,x + \frac{\partial}{\partial x}\right)\psi_0(x) = 0\,, \tag{6.10}$$

where the last equality uses $\psi_0(x) = C\,\exp\left(-\frac{1}{2}\,m\omega\,x^2\right)$ where C is a normalization constant (recall $\mathcal{H}_0(x)$ is an order-0 polynomial and so is a constant).

More generally A and $A^\star$ are *ladder operators*, in the sense that they take one energy eigenstate and give the next one, with n shifted by one. That is, we now show

$$A^\star\,|n\rangle = \sqrt{n+1}\,|n+1\rangle \qquad \text{and} \qquad A\,|n\rangle = \sqrt{n}\,|n-1\rangle\,. \tag{6.11}$$

Notice that taking the inner product of the second of these with itself agrees with (6.9). Eq. (6.11) can be proven by direct application of the definitions together with the definition of the Hermite polynomials. But an easier way to show it is to instead start from the easily proven commutation relation $[H, A^\star] = HA^\star - A^\star H = \omega\,A^\star$, together with its adjoint $[H, A] = HA - AH = -\omega\,A$. This commutation relation implies

$$\begin{aligned} H\,(A^\star\,|n\rangle) &= HA^\star\,|n\rangle = (A^\star H + \omega\,A^\star)\,|n\rangle \\ &= \left[\left(n + \frac{1}{2}\right)\omega + \omega\right] A^\star\,|n\rangle = \left[(n+1) + \frac{1}{2}\right]\omega\;\;(A^\star|n\rangle)\;, \end{aligned} \tag{6.12}$$

which uses $H\,|n\rangle = \left(n + \frac{1}{2}\right)\omega\,|n\rangle$. The above manipulations show $A^\star|n\rangle$ is also an eigenstate of H with eigenvalue $\left[(n+1) + \frac{1}{2}\right]\omega$ and so $A^\star\,|n\rangle$ must be proportional to $|n+1\rangle$, as required. The proportionality constants in (6.11) then follow from the orthogonality and normalization conditions $\langle n|m\rangle = \delta_{mn}$.

6.2 Creation and annihilation operators

The reason for the above digression on harmonic oscillators is that it is very similar to the quantum system of many noninteracting particles. The similarity arises because of the observation that harmonic oscillator energy levels are equally spaced:

$$E_{n+1} - E_n = \omega \qquad \text{for any } n. \tag{6.13}$$

This is very similar to the energy difference between states containing n identical non-interacting particles all having mass (*i.e.* rest energy) M. The energy of n such particles is then simply n times M plus whatever energy, E_0, the no-particle state (or vacuum) may have: $E = E_0 + n\,M$, and so $E_{n+1} - E_n = M$ for any n.

6.2.1 Creation and annihilation for bosons

We now formalise this resemblance more explicitly. To this end suppose we consider a non-interacting particle whose single-particle states are labelled by momentum and a collection of other labels, $|\mathbf{p}\,\sigma\rangle$, where σ denotes all of the other labels (spin, charge, baryon number, and so on) required to uniquely specify a given particle state. Then the Hilbert space of ordinary single-particle quantum mechanics is the span of these basis states: $\mathfrak{H}_1$ is the set of all states of the generic form $|\psi\rangle = \sum c(\mathbf{p},\sigma)\,|\mathbf{p}\,\sigma\rangle$ for some complex coefficients $c(\mathbf{p},\sigma)$, with the 'sum' running over all allowed values for $\mathbf{p}$ and σ.

The Hilbert space of quantum field theory is *much* larger than just the space of states in $\mathfrak{H}_1$. For instance, it also includes the space of no-particle states, spanned by the single state $|0\rangle$, so $\mathfrak{H}_0 = \{|0\rangle\}$. It also contains the space of all possible two-particle states: $\mathfrak{H}_2$, spanned by all possible states of the form $|\mathbf{p}_1\,\sigma_1\,,\mathbf{p}_2\,\sigma_2\rangle$, as well as $\mathfrak{H}_3$, spanned by all three-particle states, and so on for $\mathfrak{H}_n$ for all integers $n \geq 0$. In general it is useful to use the 'occupation number' basis, where we label a given state by the quantum number of the single-particle states that are occupied, as well as with the number of particles occupying the state. For instance

$$|(\mathbf{p}_1,\sigma_1)_{n_1};\cdots;(\mathbf{p}_r,\sigma_r)_{n_r}\rangle\,, \tag{6.14}$$

represents a state in which (for each $j = 1, 2, ..., r$) the single-particle state, $|\mathbf{p}_j\,\sigma_j\rangle$, is occupied by n_j particles, so the total number of particles present is

$$N = \sum_{j=1}^{r} n_j\,. \tag{6.15}$$

Using the operator $A^\star$ for the harmonic oscillator as a guide, we define the *creation operator* $a^\star_{p\sigma}$ as the operator that adds one particle with quantum numbers $\mathbf{p}\,\sigma$ to any given state. That is, when acting on an N-particle state we have

$$a^\star_{p\sigma}|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle = |\mathbf{p}\,\sigma;(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle\,, \tag{6.16}$$

if $\mathbf{p}\,\sigma$ is not already present, while

$$a^\star_{p\sigma}|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle = \sqrt{n_j+1}\;|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_j\,\sigma_j)_{n_j+1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle \tag{6.17}$$

if $\mathbf{p}\,\sigma = \mathbf{q}_j\,\zeta_j$ for one of the particles already present (whose label is j).

Similarly, the *annihilation operator* $a_{p\sigma}$ is defined to remove one particle with quantum number $\mathbf{p}\,\sigma$ if such a particle is present, and to give zero if no such particle is present. That is

$$a_{p\sigma}|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle = \sum_{j=1}^{r}\delta^3(\mathbf{p}-\mathbf{q}_j)\delta_{\sigma\zeta_j}\sqrt{n_j}\;|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_j\,\sigma_j)_{n_j-1};(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle\,. \tag{6.18}$$

In particular, acting on zero- and single-particle states the above specializes to

$$a_{p\sigma}|0\rangle = 0\,,\quad a^\star_{p\sigma}|0\rangle = |\mathbf{p}\,\sigma\rangle\quad\text{and}\quad a_{p\sigma}|\mathbf{q}\,\zeta\rangle = \delta^3(\mathbf{p}-\mathbf{q})\,\delta_{\sigma\zeta}|0\rangle\,, \tag{6.19}$$

while any multiple-particle state can be regarded as being obtained by applying an appropriate combination of creation operators to the vacuum:

$$|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle \propto \left(a^\star_{q_1\zeta_1}\right)^{n_1}\cdots\left(a^\star_{q_r\zeta_r}\right)^{n_r}|0\rangle\,. \tag{6.20}$$

Repeated application of these operators shows (not surprisingly) that they satisfy the same commutation relations as did[34] A and $A^\star$:

$$\left[a_{p\sigma},a^\star_{q\zeta}\right] = a_{p\sigma}a^\star_{q\zeta} - a^\star_{q\zeta}a_{p\sigma} = \delta^3(\mathbf{p}-\mathbf{q})\,\delta_{\sigma\zeta}\,. \tag{6.21}$$

Notice that this algebra only applies to bosons since only bosons can have multiply occupied states. Because a multi-particle bose state must be completely symmetric under particle interchange (this is the definition of a boson), $|\mathbf{q}_1\,\zeta_1;\mathbf{q}_2\,\zeta_2\rangle = |\mathbf{q}_2\,\zeta_2;\mathbf{q}_1\,\zeta_1\rangle$, and because $|\mathbf{q}_1\,\zeta_1;\mathbf{q}_2\,\zeta_2\rangle = a^\star_{q_1\zeta_1}a^\star_{q_2\zeta_2}|0\rangle$, it follows that we can choose $a_{p\sigma}a_{q\zeta} = a_{q\zeta}a_{p\sigma}$ and so

$$\left[a_{p\sigma},a_{q\zeta}\right] = a_{p\sigma}a_{q\zeta} - a_{q\zeta}a_{p\sigma} = 0\,. \tag{6.22}$$

Just as is true for the harmonic oscillator, the algebra defined by (6.21) and (6.22) ensures that the operator $N_{p\sigma} = a^\star_{p\sigma}a_{p\sigma}$ is diagonal in the occupation number representation. Keeping track of the density of states associated with the switch from discrete to continuum normalization, its eigenvalues count the number of particles in the following precise sense:

$$a^\star_{p\sigma}a_{p\sigma}|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle = \sum_{j=1}^{r}n_j\;\delta^3(\mathbf{p}-\mathbf{q}_j)\delta_{\sigma\zeta_j}\;|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle\,, \tag{6.23}$$

and so it is the operator $N = \sum_\sigma\int \mathrm{d}^3\mathbf{p}\;a^\star_{p\sigma}a_{p\sigma}$ that counts the number of particles present in a state:

$$\begin{aligned} N|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle &= \sum_\sigma\int \mathrm{d}^3\mathbf{p}\;a^\star_{p\sigma}a_{p\sigma}|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle \\ &= \sum_{j=1}^{r}n_j\;|(\mathbf{q}_1\,\zeta_1)_{n_1};\ldots;(\mathbf{q}_r\,\zeta_r)_{n_r}\rangle\,. \end{aligned} \tag{6.24}$$

[34] This is the same up to normalization, since we normalize momentum eigenstates different than we do harmonic oscillator states.

Consequently the Hamiltonian for free particles can be written

$$H_{\text{free}} = E_0 + \sum_\sigma \int \mathrm{d}^3\mathbf{p}\, \varepsilon(\mathbf{p}\,\sigma)\, a^\star_{p\sigma} a_{p\sigma}\,, \tag{6.25}$$

with $\varepsilon(\mathbf{p}\,\sigma)$ the single-particle energy for a state labelled by $(\mathbf{p}\,\sigma)$. For relativistic systems the single-particle energy is $\varepsilon(\mathbf{p}\,\sigma) = \sqrt{\mathbf{p}^2 + M^2(\sigma)}$, but for nonrelativistic systems it can be more complicated. The above manipulations show that H does what it should do:

$$H_{\text{free}}|\mathbf{q}_1\,\zeta_1\,n_1;\ldots;\mathbf{q}_r\,\zeta_r\,n_r\rangle = E_{\text{free}}\,(\mathbf{q}_1\,\zeta_1\,n_1;\ldots;\mathbf{q}_r\,\zeta_r\,n_r)\,|\mathbf{q}_1\,\zeta_1\,n_1;\ldots;\mathbf{q}_r\,\zeta_r\,n_r\rangle\,, \tag{6.26}$$

with

$$E_{\text{free}}\,(\mathbf{q}_1\,\zeta_1\,n_1;\ldots;\mathbf{q}_r\,\zeta_r\,n_r) = E_0 + \sum_{j=1}^{r} n_j\,\varepsilon(\mathbf{q}_j\,\sigma_j)\,. \tag{6.27}$$

In particular, the ground state is the state with the lowest-energy eigenvalue which — assuming[35] $\varepsilon(\mathbf{p}\,\sigma) > 0$ — is given by the no-particle state, $|0\rangle$, with

$$H_{\text{free}}\,|0\rangle = E_0\,|0\rangle\,. \tag{6.28}$$

Usually this *vacuum energy* is not measurable in experiments in the laboratory, because such measurements usually are sensitive only to energy differences in which E_0 cancels out. The vacuum energy density can be measured, however, through its gravitational effects since gravity responds to all energies, regardless of their origin. The evidence for the existence of *Dark Energy* in cosmology can be interpreted as the detection of the gravitational influence of the vacuum energy.

6.2.2 Creation and annihilation for fermions

As mentioned earlier, the above discussion necessarily involves particles that can multiply occupy a state, since nothing stops applying a creation operator $a^\star_{p\sigma}$ as often as one wishes to a state like $|\mathbf{p}\,\sigma\rangle$ already containing the particle in question. Consequently the formalism as described so far necessarily only applies to bosons.

How do creation and annihilation operators for fermions differ from the previous discussion? Since a fermionic state is either occupied or not, it is essentially a two-level system, rather than the infinite tower of states described above. Suppose, then, we denote the unoccupied and occupied states as follows

$$|0\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{(unoccupied)} \qquad \text{and} \qquad |1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{(occupied)}\,. \tag{6.29}$$

[35] Notice that if $\varepsilon(\mathbf{p}\,\sigma) < 0$ for any $(\mathbf{p}\,\sigma)$ then the spectrum of H_{free} is not bounded from below, since the energy can be lowered arbitrarily far just by multiply occupying any negative-energy particle states. For this reason systems with negative-energy single-particle states are usually regarded as being sick.

In this case the creation and annihilation operators are defined by the four conditions $a|0\rangle = 0$, $a|1\rangle = |0\rangle$, $a^\star|0\rangle = |1\rangle$ and $a^\star|1\rangle = 0$. This corresponds to the following explicit two-by-two matrices

$$a = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad \text{and} \qquad a^\star = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} . \tag{6.30}$$

Explicit matrix multiplication then shows that a (and $a^\star$) is *nilpotent* — that is, $a^2 = 0$ and $(a^\star)^2 = 0$ — and furthermore that a and $a^\star$ satisfy the *anticommutator* relation

$$\Big\{a^\star, a\Big\} := a^\star\, a + a\, a^\star = 1\,. \tag{6.31}$$

Furthermore direct multiplication also shows that

$$a^\star a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} , \tag{6.32}$$

and so

$$a^\star a|n\rangle = n|n\rangle \qquad (\text{for } n = 0, 1)\,. \tag{6.33}$$

Fermi statistics also determines how the creation and destruction operators for different particles or different momenta commute or anticommute. That is, fermionic states are antisymmetric under particle interchange, $|\mathbf{q}_1\,\sigma_1; \mathbf{q}_2\,\sigma_2\rangle = -|\mathbf{q}_2\,\sigma_2; \mathbf{q}_1\,\sigma_1\rangle$ (and similarly for states with more particles). Consequently, because $|\mathbf{q}_1\,\zeta_1; \mathbf{q}_2\,\zeta_2\rangle = a^\star_{q_1\zeta_1} a^\star_{q_2\zeta_2}|0\rangle$ we impose the following anticommutation relations for fermionic operators

$$\Big\{a^\star_{q_1\zeta_1}, a_{q_2\zeta_2}\Big\} := a^\star_{q_1\zeta_1}\, a_{q_2\zeta_2} + a_{q_2\zeta_2}\, a^\star_{q_1\zeta_1} = \delta^3(\mathbf{q}_1 - \mathbf{q}_2)\, \delta_{\zeta_1\zeta_2}\,, \tag{6.34}$$

and

$$\Big\{a_{q_1\zeta_1}, a_{q_2\zeta_2}\Big\} := a_{q_1\zeta_1}\, a_{q_2\zeta_2} + a_{q_2\zeta_2}\, a_{q_1\zeta_1} = 0\,. \tag{6.35}$$

The upshot is that the Hamiltonian for free fermions can also be written in precisely the same way as for bosons:

$$H_{\text{free}} = E_0 + \sum_\sigma \int \mathrm{d}^3\mathbf{p}\; \varepsilon(\mathbf{p}\,\sigma)\, a^\star_{p\sigma} a_{p\sigma}\,, \tag{6.36}$$

with $\varepsilon(\mathbf{p}\,\sigma)$ the single-particle energy for a state labelled by $(\mathbf{p}\,\sigma)$. This again does what it should do:

$$H_{\text{free}}|\mathbf{q}_1\,\zeta_1\,n_1; \ldots; \mathbf{q}_r\,\zeta_r\,n_r\rangle = E_{\text{free}}\,(\mathbf{q}_1\,\zeta_1\,n_1; \ldots; \mathbf{q}_r\,\zeta_r\,n_r)\,|\mathbf{q}_1\,\zeta_1\,n_1; \ldots; \mathbf{q}_r\,\zeta_r\,n_r\rangle\,, \tag{6.37}$$

with

$$E_{\text{free}}\,(\mathbf{q}_1\,\zeta_1\,n_1; \ldots; \mathbf{q}_r\,\zeta_r\,n_r) = E_0 + \sum_{j=1}^{r} n_j\,\varepsilon(\mathbf{q}_j\,\sigma_j)\,, \tag{6.38}$$

with the only new feature (relative to bosons) being that relations (6.34) and (6.35) replace (6.21) and (6.22) in order to ensure that the occupation numbers, n_j, always equal zero or one.

6.3 Interactions and fields

The previous sections provide a description of noninteracting particles and their energies. It also provides a language in terms of which to describe interactions that can change the number of particles.

6.3.1 Interactions

To see how this works imagine a world containing only two kinds of particles: a heavy spinless boson, h, with mass M and a lighter spin-half fermion, f, with mass $m < \frac{1}{2}M$. We assume relativistic single-particle dispersion relations, so

$$\epsilon(\mathbf{p}) = \sqrt{\mathbf{p}^2 + m^2} \quad \text{and} \quad \omega(\mathbf{k}) = \sqrt{\mathbf{k}^2 + M^2}\,, \tag{6.39}$$

for light and heavy particles respectively.

Denote the annihilation operator for the boson by a_k and the annihilation operator for the fermion by $c_{p\sigma}$, where $\sigma = \pm\frac{1}{2}$ denotes the state's 3rd component of spin. Because the heavy state is a boson and the light state is a fermion the creation and annihilation operators satisfy

$$\begin{aligned} \Big[a_k\,,a_q\Big] &= 0\,, \qquad \Big[a_k\,,a_q^\star\Big] = \delta^3(\mathbf{k}-\mathbf{q}) \\ \text{and} \qquad \Big\{c_{k\sigma}\,,c_{q\zeta}\Big\} &= 0\,, \qquad \Big\{c_{k\sigma}^\star\,,c_{q\zeta}\Big\} = \delta^3(\mathbf{k}-\mathbf{q})\,\delta_{\sigma\zeta}\,. \end{aligned} \tag{6.40}$$

The free Hamiltonian for such a system is then given in terms of these operators by

$$H_{\text{free}} = E_0 + \int \mathrm{d}^3\mathbf{p} \left[a_p^\star a_p\, \omega(\mathbf{p}) + \sum_{\sigma=\pm\frac{1}{2}} c_{p\sigma}^\star c_{p\sigma}\, \epsilon(\mathbf{p}) \right], \tag{6.41}$$

We wish to describe an interaction, H_{int}, that allows the heavy boson to decay into a pair of light fermions: $h \to ff$. To describe the decay this should have a nonzero matrix element of the form $\langle \mathbf{p}\,\sigma;\mathbf{q}\,\zeta|H_{\text{int}}|\mathbf{k}\rangle$ where the single-particle state on the right is $|\mathbf{k}\rangle = a_k^\star\,|0\rangle$ and the two particle state on the left is

$$\langle \mathbf{p}\,\sigma;\mathbf{q}\,\zeta| = \Big(|\mathbf{p}\,\sigma;\mathbf{q}\,\zeta\rangle\Big)^\star = \Big(c_{p\sigma}^\star c_{q\zeta}^\star|0\rangle\Big)^\star = \langle 0|\,c_{q\zeta}\,c_{p\sigma}\,. \tag{6.42}$$

An interaction that could describe this decay is, for example,

$$\begin{aligned} H_{\text{int}} = G \sum_{\sigma=-\zeta} \int \frac{\mathrm{d}^3\mathbf{p}}{\sqrt{(2\pi)^3 2\epsilon(\mathbf{p})}} \frac{\mathrm{d}^3\mathbf{q}}{\sqrt{(2\pi)^3 2\epsilon(\mathbf{q})}} \frac{\mathrm{d}^3\mathbf{k}}{\sqrt{(2\pi)^3 2\omega(\mathbf{k})}} \\ \times \Big[a_k\, c_{p\sigma}^\star c_{q\zeta}^\star + a_k^\star\, c_{q\zeta}\, c_{p\sigma}\Big] (2\pi)^3\delta^3(\mathbf{k}-\mathbf{q}-\mathbf{p})\,, \end{aligned} \tag{6.43}$$

where G is a real 'coupling constant' and the second term in the square bracket is the hermitian conjugate of the first one, so that H_{int} is hermitian. Applying the rules for how creation and

annihilation operators act on particle states shows that this interaction Hamiltonian has the matrix element

$$\langle \mathbf{p}\,\sigma; \mathbf{q}\,\zeta | H_{\rm int} | \mathbf{k} \rangle = \frac{G}{\sqrt{8(2\pi)^9 \epsilon(\mathbf{p})\,\epsilon(\mathbf{q})\,\omega(\mathbf{k})}}\,(2\pi)^3 \delta^3(\mathbf{k}-\mathbf{q}-\mathbf{p})\,\delta_{\sigma,-\zeta}\,. \tag{6.44}$$

Applying the standard rules of time-dependent perturbation theory in the interaction picture, as might be seen in a quantum mechanics class, shows that the transition amplitude for the process $h \to ff$ is then (to lowest order in the coupling constant G) given by

$$\begin{aligned} \mathcal{T}[h(\mathbf{p}) \to f(\mathbf{p}\,\sigma)\,f(\mathbf{q}\,\zeta)] &= -2\pi i\,\langle \mathbf{p}\,\sigma; \mathbf{q}\,\zeta | H_{\rm int} | \mathbf{k} \rangle\;\delta[\omega(\mathbf{k}) - \epsilon(\mathbf{p}) - \epsilon(\mathbf{q})] \\ &= -\frac{i\mathcal{A}}{\sqrt{8(2\pi)^9 \epsilon(\mathbf{p})\,\epsilon(\mathbf{q})\,\omega(\mathbf{k})}}\,(2\pi)^4 \delta^4(k-p-q)\,, \end{aligned} \tag{6.45}$$

where the invariant amplitude for this reaction is

$$\mathcal{A} = G\,\delta_{\sigma,-\zeta}\,. \tag{6.46}$$

Standard manipulations then show that the differential reaction rate is

$$\mathrm{d}\Gamma[h(\mathbf{k}) \to f(\mathbf{p}\,\sigma)\,f(\mathbf{q}\,\zeta)] = \frac{|\mathcal{A}|^2}{2\omega(\mathbf{k})}\,(2\pi)^4 \delta^4(k-p-q)\,\frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3 2\epsilon(\mathbf{p})}\,\frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^3 2\epsilon(\mathbf{q})}\,, \tag{6.47}$$

and so comparing with (2.29) shows that the invariant decay rate, $\mathcal{M}(h \to ff)$, as we have defined it in earlier sections is given in terms of $\mathcal{A}(h \to ff)$ by

$$\mathcal{M} = |\mathcal{A}|^2 = |G|^2\,\delta_{\sigma,-\zeta}\,. \tag{6.48}$$

It is calculations such as these that allow one to compute the invariant rates used in the exercises in previous sections given proposals for the interaction $H_{\rm int}$. For instance, in the weak interactions the term responsible for a decay like $\mu(p\,\sigma) \to e(q\,\zeta)\nu_\mu(k\,\xi)\overline{\nu}_e(r\,\upsilon)$ in $H_{\rm int}$ is proportional to something of the schematic form $G_{\scriptscriptstyle F}\,a_{p\,\sigma} b^\star_{q\,\zeta} c^\star_{k\,\xi} d^\star_{r\upsilon}$ where a, b, c and d are the appropriate annihilation operators for muons, electrons and the two types of neutrinos, with a proportionality factor that requires a more detailed treatment of the relativistic particle spins.

6.3.2 Fields

There is one feature about interactions that the above discussion makes obscure: the *locality* of interactions. That is, we expect that if systems that are sufficiently far apart from one another at a given time and start off in uncorrelated states, then their evolution should preserve their lack of correlation. Since probabilities factorize for uncorrelated systems this means we should expect that the time evolution operator, $U(t,t_0) \propto \prod_x U_x(t,t_0)$, should come to us as a product of independent evolution at different spatial points, x. But because

$U(t,t_0) = \exp[-iH(t-t_0)]$ this means that the system's Hamiltonian should come to us as a sum of independent terms for each spacetime point:

$$H = \int \mathrm{d}^3\mathbf{x}\, \mathcal{H}(\mathbf{x})\,, \tag{6.49}$$

for some energy density, $\mathcal{H}(\mathbf{x})$.

The natural variable on which $\mathcal{H}(\mathbf{x})$ should depend directly is then the combination of $a_{p\sigma}$ and $a^\star_{p\sigma}$ that corresponds to a definite position: the position-space field

$$\psi_\sigma(\mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^{3/2}}\, a_{p\sigma}\, e^{ipx} = \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^{3/2}}\, a_{p\sigma}\, e^{-i\varepsilon(\mathbf{p})t + i\mathbf{p}\cdot\mathbf{x}}\,, \tag{6.50}$$

and its adjoint

$$\psi^\star_\sigma(\mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^{3/2}}\, a^\star_{p\sigma}\, e^{-ipx} = \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^{3/2}}\, a^\star_{p\sigma}\, e^{i\varepsilon(\mathbf{p})t - i\mathbf{p}\cdot\mathbf{x}}\,. \tag{6.51}$$

For instance, if for non-relativistic particles we have the Schrödinger Hamiltonian,

$$\mathcal{H}(\mathbf{x}) = \sum_\sigma \psi^\star_\sigma(\mathbf{x}) \left[-\frac{\nabla^2}{2m}\right] \psi_\sigma(\mathbf{x})\,, \tag{6.52}$$

then substituting (6.50) and (6.51) then gives the following free-particle Hamiltonian

$$\begin{aligned} H = \int \mathrm{d}^3\mathbf{x}\, \mathcal{H}(\mathbf{x}) &= \sum_\sigma \int \mathrm{d}^3\mathbf{x} \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^{3/2}}\, a^\star_{p\sigma} \left[\frac{\mathbf{q}^2}{2m}\right] a_{q\sigma}\, e^{i(q-p)x} \\ &= \sum_\sigma \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^{3/2}} \int \frac{\mathrm{d}^3\mathbf{q}}{(2\pi)^{3/2}}\, a^\star_{p\sigma} \left[\frac{\mathbf{q}^2}{2m}\right] a_{q\sigma}\, (2\pi)^3 \delta^3(\mathbf{p}-\mathbf{q}) \\ &= \sum_\sigma \int \mathrm{d}^3\mathbf{p}\, \varepsilon(\mathbf{p})\, a^\star_{p\sigma} a_{p\sigma}\,, \end{aligned} \tag{6.53}$$

with single-particle energy

$$\varepsilon(\mathbf{p}) = \frac{\mathbf{p}^2}{2m}\,. \tag{6.54}$$

More generally, suppose

$$\mathcal{H}(\mathbf{x}) = \sum_\sigma \psi^\star_\sigma(\mathbf{x}) \left[-\frac{1}{2m}\nabla^2 + V(\mathbf{x})\right] \psi_\sigma(\mathbf{x})\,, \tag{6.55}$$

and suppose the corresponding time-independent Schrödinger equation has solutions $u_n(\mathbf{x})$,

$$\left[-\frac{1}{2m}\nabla^2 + V(\mathbf{x})\right] u_n(\mathbf{x}) = \varepsilon_n\, u_n(\mathbf{x})\,, \tag{6.56}$$

for some energy eigenvalues, ε_n. The Hamiltonian can be written in its diagonalized form in this case using the fields

$$\psi(\mathbf{x}) = \sum_n a_n\, u_n(\mathbf{x}) \quad \text{and} \quad \psi^\star(\mathbf{x}) = \sum_n a^\star_n\, u^*_n(\mathbf{x})\,, \tag{6.57}$$

since this, when substituted into (6.55) gives

$$\begin{aligned} H = \int \mathrm{d}^3\mathbf{x}\, \mathcal{H}(\mathbf{x}) &= \sum_{nm} \int \mathrm{d}^3\mathbf{x}\, a_n^\star u_n^*(\mathbf{x}) \left[-\frac{1}{2m}\nabla^2 + V(\mathbf{x}) \right] a_m\, u_m(\mathbf{x}) \\ &= \sum_{nm} \varepsilon_m a_n^\star a_m \int \mathrm{d}^3\mathbf{x}\, u_n^*(\mathbf{x})\, u_m(\mathbf{x}) \\ &= \sum_n \varepsilon_n a_n^\star a_n \,, \end{aligned} \tag{6.58}$$

where the last equality uses the orthonormality of the wave-functions $u_n(\mathbf{x})$:

$$\int \mathrm{d}^3\mathbf{x}\, u_n^*(\mathbf{x})\, u_m(\mathbf{x}) = \delta_{mn} \,. \tag{6.59}$$

In this case the single-particle energy is ε_n and the Hamiltonian (6.55) clearly describes a system of potentially many particles that interact with a potential but not with each other, with energy levels $E = \sum_n N_n \varepsilon_n$ (where N_n is the number of particles present prepared in state n).

A possible interaction term for this kind of system that does not change the number of particles might be written in position space by

$$H_{\text{int}} = \int \mathrm{d}^3\mathbf{x}\, \mathrm{d}^3\mathbf{y}\; \psi^\star(\mathbf{x})\psi(\mathbf{x})U(\mathbf{x}-\mathbf{y})\psi^\star(\mathbf{y})\psi(\mathbf{y}) \,, \tag{6.60}$$

while an interaction that describes the emission and absorption of a boson destroyed by the operator b_k might be written

$$H_{\text{int}} = \int \mathrm{d}^3\mathbf{x}\; \psi^\star(\mathbf{x})\, \psi(\mathbf{x}) \Big[g\, \phi(\mathbf{x}) + g^* \phi^\star(\mathbf{x}) \Big] \,, \tag{6.61}$$

where g is a coupling constant, $\phi(\mathbf{x}) = \sum_k b_k\, e^{ikx}$ and so on.

6.4 Relativistic quantum field theory

The relation between the field in position space and the creation and annihilation operators takes a different form in relativistic field theories. Rather than expressions like (6.50) and (6.51), in relativistic theories one instead always finds the position-space field is given by expressions like

$$\psi(\mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{p}}{\sqrt{(2\pi)^3 2\varepsilon(\mathbf{p})}} \left[a_p\, e^{ipx} + \bar{a}_p^\star\, e^{-ipx} \right] , \tag{6.62}$$

where a_p destroys the particle of interest and $\bar{a}_p$ destroys its antiparticle.[36]

[36] One way to think about the need for antiparticles to enter into fields this way comes from the problem of reconciling *relativity of simultaneity* with the Heisenberg uncertainty principle. Relativity of simultaneity states that different observers can disagree on the ordering of two events, A and B, in time provided A and

The antiparticle term in (6.62) turns out to be necessary in order to ensure that the field and its adjoint commute (or anticommute, for fermions) for space-like separations, and this is in turn ultimately required in order for $\mathcal{H}(x)$ to commute with $\mathcal{H}(y)$ when x^μ and y^μ are space-like separated. And this commutation of $\mathcal{H}(x)$ with itself at spacelike-separated points is required in order for interactions to preserve Lorentz-invariance. This necessity can be seen, for instance, from the ubiquitous appearance of time-ordered correlation functions like $\langle 0|T[\mathcal{H}(x)\,\mathcal{H}(y)]|0\rangle$ in perturbative calculations of scattering in quantum mechanics. Here the symbol $T[\mathcal{H}(x)\,\mathcal{H}(y)]$ denotes the time-ordered product: equal to $\mathcal{H}(x)\,\mathcal{H}(y)$ if $y^0 > x^0$ (*i.e.* the event x^μ is later than y^μ), but equal to $\mathcal{H}(y)\,\mathcal{H}(x)$ if $x^0 > y^0$ (*i.e.* when x^μ is earlier than y^μ). But the relativity of simultaneity implies the relative size of x^0 and y^0 is ambiguous when x^μ and y^μ are space-like separated, because different observers can disagree on the ordering in time of space-like separated events. So such time-orderings of $\mathcal{H}$ can only be unambiguous in a relativistic theory if $\mathcal{H}(x)$ commutes with $\mathcal{H}(y)$ at space-like separation.

It is this requirement that fields and their adjoints commute/anticommute at spacelike separations (sometimes called *micro-causality*) that is ultimately at the origin of the need for antiparticles. It is also at the root of a number of other very general consequences of the reconciliation of special relativity and quantum mechanics, whose explanation goes beyond the scope of these notes but which are listed here for completeness.

- **Existence of antiparticles:** For every type of particle, p, there also exists an antiparticle, $\overline{p}$, which in some circumstances (for particles that carry no conserved charges — see below) can be the same as the particle.

- **Inevitability of particle-number changing interactions:** Particles and antiparticles enter interactions only through the fields, which schematically depend on particle and antiparticle creation and annihilation operators only through the combination $\psi \sim a + \overline{a}^\star$. Consequently while ψ destroys particles its antiparticle part creates antiparticles, and they enter with the same relative strength. This implies that any interactions

B are space-like separated: $s^2(A,B) = (\Delta\mathbf{x})^2 - (\Delta t)^2 > 0$. In relativity this inability to agree on ordering in time does not preclude predicting the future from the past (which, after all, is the goal of science) because no information can travel faster than light. A cannot influence B (or vice versa) because doing so requires something to move faster than the speed of light. But this argument breaks down in quantum mechanics, because if you know you are precisely at A then you must be arbitrarily uncertain about your momentum and so there can actually be a nonzero quantum amplitude for you to move faster than light and so have A influence B, say. But since some observers see B to be *earlier* than A they must also have a story to tell. In relativistic quantum field theory the story that makes everything consistent is that if one observer sees A earlier than B and sees a particle carry charge, Q, and energy, E, from A to B, then for an observer with B earlier than A there must be another particle (which we call the antiparticle) that carries charge $-Q$ and energy E from B to A, with exactly the same amplitude. This is ultimately why antiparticles exist, and why their charges are precisely opposite (and their mass precisely equal) to those of the particles to which they correspond, and why they enter into interactions with precisely the same strength.

built from $\psi(\mathbf{x})$ and $\psi^\star(\mathbf{x})$ necessarily must change particle number, so it is not really possible to have Lorentz-invariant interactions that preserve the number of particles at all energies.

- **Crossing Symmetry:** Because the particle and antiparticle parts enter into ψ with the same relative strength, the amplitude for any process with particle p in the final (initial) state is precisely the same as that for the process with its antiparticle $\overline{p}$ in the initial (final) state. For example the amplitude for $n \to p + e^- + \overline{\nu}_e$ is identical to the amplitude for $e^+ + n \to p + \overline{\nu}_e$ and for $\nu_e + n \to p + e^-$ and for $n + \overline{p} \to e^- + \overline{\nu}_e$ and so on. Of course the phase space for these various reactions need not be the same and this can cause differences in their overall rates at any given energy.

- **CPT Symmetry:** Because a and $\overline{a}$ appear only through the schematic combination $a + \overline{a}^\star$ within ψ, particles and antiparticles must have precisely opposite charges for any conserved additive charge (like electric charge or lepton number): *i.e.* $Q(p) = -Q(\overline{p})$ and $L(p) = -L(\overline{p})$. They must also have precisely equal masses: $M(p) = M(\overline{p})$. These requirements require the particle and antiparticle to be different from one another whenever either carries a nonzero conserved charge. The precise equality of properties can be formalized by the statement that any local relativistic (and unitary) quantum system has a symmetry called CPT, consisting of simultaneous interchange of particle and antiparticle (C: $a \leftrightarrow \overline{a}$), parity (or reflection of all spatial coordinates, P: $\mathbf{x} \leftrightarrow -\mathbf{x}$) and time-reversal (T: $t \leftrightarrow -t$).

- **Spin-Statistics connection:** Having $\psi \sim a + \overline{a}^\star$ actually ensures interactions commute for spacelike separations, $[\mathcal{H}(\mathbf{x}, t), \mathcal{H}(\mathbf{y}, t)] = 0$, only if all integer-spin ($s = 0, 1, 2, \dots$) particles satisfy Bose statistics and all half-odd-integer spin ($s = \frac{1}{2}, \frac{3}{2}, \dots$) particles satisfy Fermi statistics.

6.4.1 Quantum electrodynamics

The poster child of a relativistic quantum field theory is Quantum Electrodynamics, in which it is the electromagnetic field that gets expanded in terms of creation and annihilation operators for photons, as in (6.62). That is, consider electric and magnetic fields describing an electromagnetic wave, which can be written in terms of the vector potential $\mathbf{A}$ as

$$\mathbf{E} = \frac{\partial \mathbf{A}}{\partial t} \qquad \text{and} \qquad \mathbf{B} = \nabla \times \mathbf{A}\,, \tag{6.63}$$

where

$$\mathbf{A}(\mathbf{x}) = \sum_{\lambda = \pm 1} \int \frac{\mathrm{d}^3\mathbf{k}}{\sqrt{(2\pi)^3 2\,\omega(\mathbf{k})}} \left[\mathbf{e}(\mathbf{k}\,\lambda)\, a_{k\lambda}\, e^{ikx} + \mathbf{e}^*(\mathbf{k}\,\lambda)\, a^\star_{k\lambda}\, e^{-ikx} \right], \tag{6.64}$$

where $\mathbf{e}(\mathbf{k}, \lambda)$ denotes the polarization vector of the photon with momentum $\mathbf{k}$ and helicity $\lambda = \pm 1$, and $kx = -\omega(\mathbf{k})\, t + \mathbf{k} \cdot \mathbf{x}$ where the photon energy is $\omega(\mathbf{k}) = |\mathbf{k}|$.

In this case the free Hamiltonian for noninteracting photons is simply the usual expression for the energy density of the field in terms of electric and magnetic fields,

$$\begin{aligned} H_{\rm free} &= \int \mathrm{d}^3\mathbf{x}\,\Big[\rho_0 + \frac{1}{2}\Big(\mathbf{E}^2 + \mathbf{B}^2\Big)\Big] \\ &= \int \mathrm{d}^3\mathbf{x}\,\rho_0 + \frac{1}{2}\sum_{\lambda=\pm1}\int \mathrm{d}^3\mathbf{k}\Big[a^\star_{k\lambda}a_{k\lambda} + a_{k\lambda}a^\star_{k\lambda}\Big]\omega(\mathbf{k}) \\ &= E_0 + \sum_{\lambda=\pm1}\int \mathrm{d}^3\mathbf{k}\; a^\star_{k\lambda}a_{k\lambda}\;\omega(\mathbf{k})\,, \end{aligned} \tag{6.65}$$

where the second equality uses (6.63) and (6.64) and ρ_0 parameterizes the classical energy density of the vacuum. As expected, this Hamiltonian describes a collection of photons whose single-particle energies add to give an energy eigenvalue

$$E = E_0 + \sum_{\lambda}\int \mathrm{d}^3\mathbf{k}\, N_{k\lambda}\,\omega(\mathbf{k})\,, \tag{6.66}$$

when acting on a state with $N_{k\lambda}$ photons with momentum $\mathbf{k}$ and helicity λ.

Notice that the total vacuum energy (*i.e.* energy of the no-particle state), E_0, obtained here, given explicitly by

$$E_0 = \int \mathrm{d}^3\mathbf{x}\,\rho_0 + \delta^3(0)\int \mathrm{d}^3\mathbf{k}\,\frac{1}{2}\,|\mathbf{k}|\,, \tag{6.67}$$

hides many sins. In particular it diverges in two separate ways (at long and short distances), and such divergences are very common in quantum field theory. First, both terms in E_0 diverge at long distances in the infinite-volume limit. The first term does so because ρ_0 is a constant and so $\int \mathrm{d}^3\mathbf{x}\,\rho_0 = \rho_0\,V$ where $V\to\infty$ is the volume of space. The second term also diverges for long distances — *i.e.* at long wavelengths, or the 'infrared' (or IR) — because it is proportional to the momentum-space delta-function: $\delta^3(0)$. This is also an infinite-volume divergence because $\delta^3(\mathbf{k}) = (2\pi)^{-3}\int \mathrm{d}^3\mathbf{x}\; e^{-i\mathbf{x}\cdot\mathbf{k}}$ implies $\delta^3(0) = \lim_{V\to\infty} V/(2\pi)^3$, where again V is the volume of space.

Like most IR divergences the divergence with infinite V is telling us we are calculating the wrong thing. In the present instance it is telling us that it is the *energy density*, $\rho_{\rm vac} = H/V$, that is the observable quantity[37] that has a chance to be finite at infinite volume rather than the total energy.

But this is not the end of the divergence story because there is also a problem at short distances, or the 'ultraviolet' (or UV), since

$$\lim_{V\to\infty}\frac{E_0}{V} = \rho_{\rm vac} = \rho_0 + \frac{1}{2}\int\frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,|\mathbf{k}|\,. \tag{6.68}$$

[37] Indeed, the vacuum energy density can be observed – by the way it gravitates, for instance.

This is infinite because $\int \mathrm{d}^3\mathbf{k}\,|\mathbf{k}| \propto \int \mathrm{d}k\; k^3$ diverges quartically as the upper integration limit goes to infinity. UV divergences such as these are usually *renormalized*: since nothing physical depends separately on ρ_0 and the divergent integral in (6.68) we can imagine that the bare parameter ρ_0 also diverges in such a way that the sum in (6.68) remains finite. This is all that is required because it is only $\rho_{\rm vac}$ (as opposed to ρ_0) that is observable, and so must be finite.

6.5 Bosons and forces

The upshot of the previous sections is that there is a creation and annihilation operator for all types of particles. So once the Hamiltonian is written in local form this also means that there is also a separate field for each type of particle, obtained by Fourier transforming the corresponding creation and annihilation operators (for the particle and its antiparticle in the schematic combination $a + \overline{a}^\star$).

But at the classical level we normally associate fields with forces, and although some particles and fields do seem to be associated with forces (such as the photon, gluon or graviton) we do not normally associate forces with all particle types (such as electrons or protons). Why is this, and what decides which particles are associated with forces?

The answer to this is that any particle satisfying Bose statistics can in principle mediate a force, although for the force to be described by a classical field usually also requires the particle mass to be quite small compared with the energies of interest. Bose statistics and low masses are required because field states begin to look like classical fields once they involve large particle occupation numbers, and both Fermi statistics and low energy block the occurrence of large occupation numbers.

For instance, the interactions of the electromagnetic field with matter usually take the form

$$H_{\rm int} = \int \mathrm{d}^3\mathbf{x}\; \mathbf{J}(\mathbf{x}) \cdot \mathbf{A}(\mathbf{x})\,, \tag{6.69}$$

where the electric current operator, $\mathbf{J}(\mathbf{x})$, is built out of the fields for electrically charged particles, like electrons. In static situations current conservation requires $\nabla \cdot \mathbf{J} = 0$. Notice in particular that $H_{\rm int}$ is linear in $\mathbf{A}$, and so is also linear in the photon creation and annihilation operators once expressed in terms of $a_{k\lambda}$ and $a^\star_{k\lambda}$. Because of this linearity $H_{\rm int}$ does not commute with photon number,

$$N = \sum_\lambda \int \mathrm{d}^3\mathbf{k}\; a^\star_{k\lambda} a_{k\lambda}\,, \tag{6.70}$$

and so when the source currents are sufficiently strong the eigenstates of the electromagnetic field can be driven not to have a definite number of photons.

When an interaction like $H_{\rm int}$ that is linear in a bosonic field happens to play an dominant role in a physical process then the system often is well-described by a *coherent state*, defined

as an eigenstate of the operator $a_{k\lambda}$ rather than N or $N_{k\lambda}$. That is, suppressing the labels $(\mathbf{k}\,\lambda)$, a coherent state, $|\alpha\rangle$, satisfies:

$$a|\alpha\rangle = \alpha|\alpha\rangle\,, \tag{6.71}$$

for some complex number α. This eigenvalue condition can be solved explicitly, leading to the following expression for $|\alpha\rangle$ in terms of the occupation-number basis, $|n\rangle$:

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}}\,|n\rangle = e^{-\frac{1}{2}|\alpha|^2}\; e^{\alpha a^\star}\;|0\rangle\,. \tag{6.72}$$

Particle number is clearly not diagonal in these states, and instead the probability of detecting n particles is given by a Poisson distribution,

$$P(n) = |\langle n|\alpha\rangle|^2 = e^{-\bar{n}}\;\frac{\bar{n}^n}{n!}\,, \tag{6.73}$$

where

$$\bar{n} := \langle n\rangle = \langle\alpha|a^\star a|\alpha\rangle = |\alpha|^2\,, \tag{6.74}$$

is the mean value for n. The variance of n in such a distribution is similarly

$$(\Delta n)^2 := \langle (n-\bar{n})^2\rangle = \langle n^2 - \bar{n}^2\rangle = |\alpha|^2\,, \tag{6.75}$$

and so $\bar{n}^2 \gg (\Delta n)^2$ whenever $\bar{n} = |\alpha| \gg 1$, showing that fluctuations become relatively small once the average occupation number becomes large.

Another way to see why $|\alpha\rangle$ behaves like a classical field when $|\alpha| \gg 1$ can be found by comparing the expectation value of $a^\star a$ with that of $a\,a^\star$, using $[a^\star, a] = 1$:

$$\langle\alpha|a^\star a|\alpha\rangle = |\alpha|^2 \qquad \text{while} \qquad \langle\alpha|a\,a^\star|\alpha\rangle = \langle\alpha|a^\star a + 1|\alpha\rangle = |\alpha|^2 + 1\,, \tag{6.76}$$

showing that inside expectation values $\langle a\,a^\star\rangle$ and $\langle a^\star a\rangle$ are approximately equal (and so a and $a^\star$ behave effectively like classical commuting variables inside expectation values) provided that $\bar{n} = |\alpha|^2 \gg 1$. Coherent states represent one example of how states involving large occupation numbers for bosons can be approximately well-described by classical fields.

To summarize: because bosons love to congregate, systems containing bosons with small masses (relative to the system energy) often evolve into states that are multiply occupied by enormous numbers of bosons. Such states are often well-described by classical fields, and it is because of this that bosons can mediate interactions between other particles. We tend therefore to associate elementary bosons (and their associated fields) with fundamental interactions.

7 The Standard Model

This section gives a brief summary of the particle content and some of the properties and puzzles of the Standard Model, which is the quantum field theory that describes all but a very few relatively recent experiments and observations. (A list of the apparent failures of the Standard Model is given in the final subsection.)

7.1 Fermions and the generation puzzle

We start with a summary of the Standard Model's 'matter content', which is equivalent to listing all of the fermions that are known and believed to be fundamental (in that there is no evidence for their having any substructure).

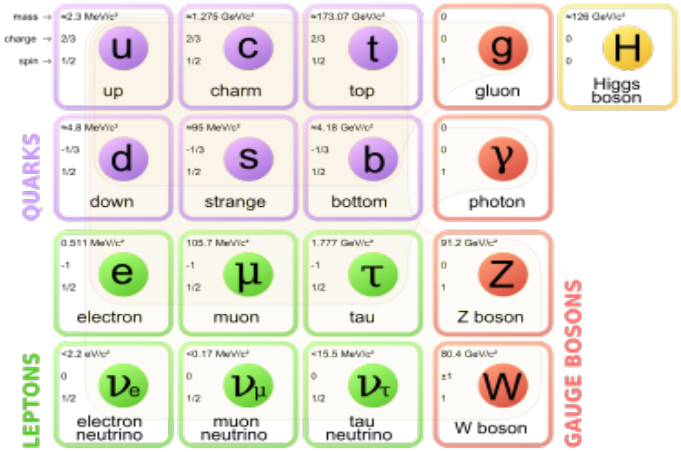


Figure 45. A table listing the particle content of the Standard Model. (Figure source: Wikipedia `https://en.wikipedia.org/wiki/Standard_Model`).

The fundamental fermions all have spin $\frac{1}{2}$ (consistent with the spin-statistics theorem of relativistic quantum mechanics) and are shown in violet and green in Figure 45. All fermions come in triplicate: there are four basic fermion types (the left-most column of 45) plus two copies of these (the next two columns) that share all of the same charge assignments of the first column, differing only in their mass.

Each column of fermions is called a *generation* and seems to be self-contained inasmuch as the theory could have been consistent if there were only one or two generations. Nobody knows why three generations appear in nature, a piece of current ignorance that is called the *Generation Puzzle*.

7.1.1 Leptons

Each generation contains two kinds of particles that do not take part in the strong interactions (more about which later), called *leptons*. Of these, the charged leptons (e, μ and τ) must differ from their antiparticles because they carry electric charge. Because they have spin-half there is a total of 4 spin states for each charged lepton: two spins each for the particle:[38] e_L^-, e_R^-, and two more for the antiparticles of these: e_R^+ and e_L^+. Plus copies of this for each of the next two generations. The masses of the charged leptons turn out to be quite different from

[38] Here we conventionally denote the two spin states as 'left-handed' and 'right-handed', by projecting their spin along the direction of their motion.

one another: $m_e \simeq 0.511$ MeV, $m_\mu \simeq 105$ MeV and $m_\tau \simeq 1.78$ GeV. As discussed above, at most one of the neutrinos can be massless, and the rest are known to have masses smaller than 1 eV or so.

For each of the neutrinos the Standard Model includes only left-handed particles (and their right-handed antiparticles), ν_{eL} and $\overline{\nu}_{eR}$ plus their generational copies, since there is no unambiguous evidence for the existence of any other spin states. Whether ν_e and $\overline{\nu}_e$ are distinct from one another or not depends on whether or not there exists a conserved charge that distinguishes them, and so hinges on whether or not it turns out that total lepton number, $L = L_e + L_\mu + L_\tau$, is conserved.

It is common to group together the particles that the weak interactions allow transitions between. This leads them to be grouped into 2-component column vectors, as in

$$\begin{pmatrix} \nu_{eL} \\ e^-_L \end{pmatrix} \qquad \begin{pmatrix} \nu_{\mu L} \\ \mu^-_L \end{pmatrix} \qquad \begin{pmatrix} \nu_{\tau L} \\ \tau^-_L \end{pmatrix} , \tag{7.1}$$

where only left-handed particles are included because all evidence is that only these (and their antiparticles) take part in weak-interaction transitions of the form $e^- \leftrightarrow \nu_e$ (or its generational counterparts). The spin-dependence of particles taking part in these interactions can be inferred from the angular distribution of the final particles in collisions and decays.

7.1.2 Quarks and the hadronic zoo

The other two kinds of fermions in each generation are quarks, which (unlike leptons) do participate in the strong interactions. There are two species of quarks in each generation: an up-type quark with charge $+\frac{2}{3}$ and a down-type one with charge $-\frac{1}{3}$. Since each of these carries charge they are distinct from their antiparticles, and because all have spin half each has two possible spin states. This leads to the independent states u_L, u_R, d_L and d_R plus their respective antiparticles $\overline{u}_R$, $\overline{u}_L$, $\overline{d}_R$ and $\overline{d}_L$ (plus copies of all of these for each of the other two generations).

The pattern of quark masses is also quite varied: $m_u \simeq 2.3$ MeV, $m_d \simeq 4.8$ MeV, $m_s \simeq 95$ MeV, $m_c \simeq 1.28$ GeV, $m_b \simeq 4.2$ GeV and $m_t \simeq 173$ GeV. There is no fundamental understanding of what determines the pattern of masses seen for the various fundamental fermions, and this lack of understanding is sometimes called the *Flavour puzzle*.

We have already seen that the up and down quarks combine under the strong interactions into a variety of states called hadrons. These come in two main types: mesons (built from quark-antiquark pairs) and baryons (built from 3-quark combinations). For up and down quarks the ground states for these bound states were the proton, neutron, Δ baryons and the π and ρ mesons. Many more combinations are possible once the full complement of six quark types are included, and of these all of those expected to be observable are seen.

It is also true for quarks that only left-handed particles seem to take part in the weak interactions that change the species of quark, and so these are often grouped together into

'doublets' (as was done for leptons):

$$\begin{pmatrix} u_L \\ d'_L \end{pmatrix} \qquad \begin{pmatrix} c_L \\ s'_L \end{pmatrix} \qquad \begin{pmatrix} t_L \\ b'_L \end{pmatrix} , \tag{7.2}$$

where the primes on the down quarks denote the following linear combinations

$$\begin{aligned} d'_L &= V_{ud}\, d_L + V_{us}\, s_L + V_{ub}\, b_L \\ s'_L &= V_{cd}\, d_L + V_{cs}\, s_L + V_{cb}\, b_L \\ b'_L &= V_{td}\, d_L + V_{ts}\, s_L + V_{tb}\, b_L \, . \end{aligned} \tag{7.3}$$

The elements of the 3-by-3 matrix V of complex coefficients can be inferred by measuring weak decays involving different species of quarks. We have already seen how V_{ud} is measured in super-allowed nuclear beta decays, such as $^{14}\mathrm{C} \to {}^{14}\mathrm{N}$, and it turns out that V_{us} can be measured in K-meson decays (where K mesons are built using s quarks) and so on.

The implications of the Standard Model for the properties of weak decays of quarks are very predictive despite the presence of the matrix V_{ij}. It is predictive because the matrix V is unitary: $V^\dagger V = V V^\dagger = I$, and many of its phases can be absorbed into five of the six quark fields.[39] So the nominally 9 complex parameters, V_{ij}, are subject to 9 real unitarity conditions plus 5 rephasing conditions and so can really be expressed in terms of 4 real parameters: the three angles of a 3-by-3 rotation, $(\theta_1, \theta_2, \theta_3)$, plus a physical phase[40] $e^{i\delta}$. There are many more than 4 observables amongst quark weak interactions, and their agreement with the Standard Model's predictions provides many nontrivial tests of the structure of V_{ij}.

7.2 Bosons and the four forces

As discussed above, fundamental bosons in the Standard Model tend to be associated with forces. Although not normally included in the Standard Model, this includes the spin-2 graviton which is the particle associated with waves in the gravitational field.

For the Standard Model proper there are a variety of bosons, each associated with a known 'force.' All but one of these has spin-1, and the exception has spin zero. Besides the photon of electromagnetism we have already encountered the 8 gluons whose exchange mediates the strong force that binds quarks into hadrons, and whose residual interactions over longer distances are what was historically called the nuclear force. Both the photon and the gluon are massless in the Standard Model (as also is the graviton).

To these familiar spin-1 particles the Standard Model adds two more: the electrically charged W^- particle (and its antiparticle W^+), and the electrically neutral Z^0 (that is its

[39] The sixth phase corresponds to a common rotation of all quarks by the same phase and is a symmetry responsible for baryon number conservation, and so cannot be used to remove a phase from V.

[40] This phase is the parameter ultimately responsible for CP-violation (*i.e.* the breaking of the symmetry obtained by combining *charge conjugation* (C) – the swapping of particles with antiparticles – with *parity* (P) – the reversing of the spatial coordinate, $\mathbf{r} \to -\mathbf{r}$) in the Standard Model.

own antiparticle, as is also true for the photon). These have similar masses: the W boson has mass $m_W \simeq 80$ GeV and the Z boson has mass $m_Z \simeq 90$ GeV. What are the forces associated with these particles?

7.2.1 Charged-current weak interactions

The 'force' associated with the W boson is the weak interaction that is responsible for β-decay and the other decays encountered to date. Within the Standard Model any element of one of the doublets, (7.1) or (7.2), can turn into the other element of the same doublet, while either emitting or absorbing a W boson. That is, within the Standard Model a decay like $d \to u + e^- + \overline{\nu}_e$ really takes place in two steps: $d \to u + W^-$ followed by the subsequent process $W^- \to e^- + \overline{\nu}_e$. Quantum mechanically the combined reaction can take place even if the mass difference $m_d - m_u \simeq 1$ MeV is too small to account for the W-boson mass, which is now known to be $m_W \simeq 80$ GeV, provided the emission and absorption of the W boson takes place sufficiently quickly and over small enough distances. If these distances and times are short enough then the uncertainty principle can allow the process to occur even though it would have been forbidden if the W had survived for a long time. When the mass of the decaying particle is much smaller than the mass of the W boson (that is, $m \ll 80$ GeV) then the Standard Model description approximately reduces to the description of the Fermi theory, with G_F calculable in terms of more fundamental parameters.

In the Standard Model the emission or absorption of a W boson is associated with a coupling parameter, g, in the same way that photon emission and absorption is associated with the electromagnetic coupling e. Since (at low energies) a W must be both emitted and absorbed in order to have a decay, we expect any decay amplitude to be proportional to g^2. Furthermore, the sense in which the Fermi theory emerges from the Standard Model is as the leading term in a Taylor expansion in powers of the decay energy divided by m_W^2. Since in the Fermi theory it is G_F that gets compared with the decay energy when determining the rate, we expect to find $G_F \propto g^2/m_W^2$ and this is indeed borne out by calculations, with the precise relation given by

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}\,. \tag{7.4}$$

Measurements of G_F and m_W thereby determine the value of g. It is conventional to express this value relative to the electromagnetic coupling in terms of an angle, θ_W, by

$$\sin^2\theta_W := \frac{e^2}{g^2} = \frac{\alpha}{\alpha_w} \simeq 0.23126(5)\,, \tag{7.5}$$

where $\alpha_w = g^2/4\pi$ is the weak fine-structure constant, analogous to $\alpha = e^2/4\pi$. Notice that α_w and α are not so different in size, and indeed the modern understanding of why the weak interactions were historically regarded as being weak is because all early experiments and decays were measured at energies much smaller than m_W, rather than because the underlying coupling, g, is particularly small.

7.2.2 Neutral-current weak interactions

At 90 GeV the Z boson mass is just slightly larger than the W boson mass, and in the Standard Model these masses are predicted to be related to one another by the (very successful) mass formula

$$m_W = m_Z \cos\theta_W \,, \tag{7.6}$$

where θ_W is the same angle as was defined in (7.5).

The Z boson is also associated with a weak interaction, called the *neutral-current* weak interaction, which is mediated when Z bosons are emitted and absorbed. But this interaction in the Standard Model is not responsible for any decays, because in the Standard Model a fermion does not change type when emitting or absorbing a Z boson. That is, while processes like $u \leftrightarrow u + Z$ or $\tau \leftrightarrow \tau + Z$ occur in the Standard Model processes like $\mu \leftrightarrow e + Z$ or $s \leftrightarrow d + Z$ are unobservably rare (given the present state of the experimental art), even though not strictly forbidden by any conservation laws.

Hypothetical flavour-changing Z processes like these are called *flavour-changing neutral currents* (or FCNCs) for historical reasons, and their observed absence is a strong piece of evidence that supports the Standard Model over many of its alternatives. For instance if the process $s \leftrightarrow d + Z$ were possible then the charged meson K^+ (consisting of a $u\overline{s}$ quark combination) could decay through the process

$$K^+ = u\,\overline{s} \to u\,\overline{d} + Z \to u\,\overline{d} + (e^+e^-) = \pi^+e^+e^- \,. \tag{7.7}$$

If the amplitude for $s \leftrightarrow d + Z$ were comparable in size to the emission/absorption processes $s \leftrightarrow s + Z$ or $d \leftrightarrow d + Z$ that do occur in the Standard Model, then it would be given by $g/\cos\theta_W = e/(\sin\theta_W \cos\theta_W)$ and so not be too different from the amplitude for emitting a photon or a W boson. Consequently a process like (7.7) would happen for an appreciable fraction of K^+ decays. Since no such decays are seen it is a big success that the Standard Model does not allow flavour-changing couplings for the Z.

An even stronger constraint on flavour-changing neutral currents comes from the neutral K mesons, of which there are two: the K^0 and $\overline{K}^0$, respectively consisting of $d\,\overline{s}$ and $s\,\overline{d}$ quark combinations. If the process $s\,\overline{d} \leftrightarrow Z$ were possible then because the Z is its own antiparticle CPT symmetry[41] implies that the process $d\,\overline{s} \leftrightarrow Z$ would also be allowed. The two of these would then allow the transition $K^0 \leftrightarrow \overline{K}^0$ through the process

$$K^0 = d\,\overline{s} \to Z \to s\,\overline{d} = \overline{K}^0 \,. \tag{7.8}$$

The limits on the level at which this process can occur are particularly strong and so again the absence of flavour-changing Z couplings is crucial to the Standard Model's success.

[41] CPT symmetry is the combination of charge conjugation (C: swapping particles with antiparticles), parity (P: reflecting $\mathbf{r} \to -\mathbf{r}$) and time-reversal (T: reflecting $t \to -t$). This combination of transformations is important because it is a theorem that every Lorentz-invariant, unitary and local quantum field theory (including in particular the Standard Model) is automatically also CPT invariant.

7.2.3 The Higgs boson

The only elementary boson in the Standard Model that is not spin-1 is the Higgs boson, which is spinless. The Higgs boson was the last Standard Model particle to be found, being discovered only in 2013 with a mass $m_h = 125$ GeV. The Higgs particle plays a special role in the Standard Model because it couples to other Standard Model particles by an amount proportional to the other particle's mass. That is, in the Standard Model a Higgs particle can be emitted by any other particle, though (just as for Z bosons) the emitting particle never changes flavour when doing so: for example $e \leftrightarrow e + h$ and $t \leftrightarrow t + h$ can happen but $t \leftrightarrow c + h$ never happens. The amplitude for Higgs emission by a particle f is proportional to m_f/v where m_f is the f particle mass and $v \sim G_F^{-1/2} = 246$ GeV is a fundamental scale in the Standard Model.

The reason the Higgs couples this way is the unusual way mass arises in the Standard Model. It turns out that the field, H, associated with the Higgs has an unusual property: it costs less energy to have the Higgs field be present than it does not to have it, so H is nonzero in the vacuum (unlike the fields for other particles), regardless of whether or not other sources are present. In the absence of the Higgs field particles in the Standard Model would be massless, and acquire nonzero masses only because of their interactions with this Higgs field in the vacuum (a process known as the *Higgs mechanism*).

7.3 Where the Standard Model fails

Although the Standard Model is an extremely successful synthesis of what we know about the structure of Nature, it gets a few things wrong and so these notes close with a brief summary of five of its known problems.

Neutrino Oscillations

The Standard Model predicts that neutrino masses vanish, and so cannot in itself describe the evidence for neutrino oscillations. These oscillations can be described by applying one of two minimal tweaks to the model, both of which amount to adding masses for neutrinos.

The simplest such tweak does not add any new degrees of freedom, and just gives masses to the three Standard Model neutrino species. This tweaked model then predicts that overall lepton number is not conserved and neutrinos are their own antiparticles (or, in the jargon, neutrinos are *Majorana* in nature). If so this would predict the existence of *neutrinoless double-β decay*, in which two β decays happen in immediate succession without the emission of two associated neutrinos, $(A, Z) \to (A, Z + 2) + 2e^-$, because the two neutrinos can now mutually annihilate. Although such decays have been sought, none has yet been seen.

Slightly more complicated variations on this theme add new particle states — so-called right-handed neutrinos — that can pair off with the Standard Model neutrinos to give them masses through their Higgs interactions, in much the same way as all other Standard Model

fermion masses arise. Such constructions can, but need not, accommodate unbroken overall lepton number, in which case neutrinoless double-β decay would never be seen. The required right-handed neutrinos turn out not to couple at all to any of the spin-one particles and so do not participate in *any* Standard Model interactions apart from the very feeble Higgs one associated with their mass. Such particles consequently can interact even more weakly than do the Standard Model neutrinos, and so are called *sterile* neutrinos. Although difficult to detect, their presence can be sought in neutrino experiments (with no evidence for them yet having arisen).

Dark Matter

Cosmologists have assembled a convincing picture of the universe around us and how it is evolving, called the Hot Big Bang (or ΛCDM) model. This model — sometimes called the Standard Model of Cosmology — very successfully describes many more observations than it has parameters, and so works well even though it is over-determined. Among the observations that are so well-described are measurements of the total average energy density of the universe, regardless of whether or not this energy is directly visible, and these observations indicate that 95% of what is out there is not contained in the Standard Model.

Some 25% of what is out there is called *Cold Dark Matter* or CDM, and seems to be some kind of matter that gravitates as would a non-relativistic species of particle. There are many separate lines of evidence pointing to CDM, including how galaxies rotate; the speed with which galaxies move within clusters of galaxies; the amount of gravitational lensing caused by clusters of galaxies and the properties of the relic Cosmic Microwave Background (CMB) radiation, left over from an earlier epoch when electrons and nuclei first bound together into atoms (after which the universe first became transparent).

We know the CDM cannot be ordinary matter (as described in the Standard Model) because the total amount of this can also be inferred from the rates with which nuclei were formed in in the early universe, as well as the measured cosmic speed of sound as inferred from the properties of the CMB. Both only allow about 5% of the total energy density to be ordinary matter, leaving the rest unexplained by any Standard Model physics.

Notice that 'Standard Model physics' here can also be taken to include using General Relativity to describe gravity, since the evidence for CDM in cosmology comes entirely from the way visible matter responds to the CDM's gravitational field. This evidence would therefore need re-examining should another theory of gravity be found to provide a better description. Attempts to tweak the Standard Model to explain CDM therefore group roughly into two types: modify gravity or invent a new very weakly interacting type of particle (perhaps the sterile neutrino mentioned above?).

Dark Energy

Earlier sections alluded to the idea that the vacuum energy density has been detected in cosmology, where it is called the Dark Energy and turns out to make up the remaining 70% of the cosmic energy budget. The nature of Dark Energy is also largely unknown, though it is known to be different from Dark Matter (and ordinary matter) because it causes the universal expansion to accelerate (which Dark Matter and ordinary matter cannot do).

Dark Energy might be consistent with the Standard Model if it is just the vacuum energy, but the problem in this case is that it is possible to compute in the Standard Model how large the vacuum energy density should be and what is observed is many orders of magnitude too small. It is as if something unknown is making the vacuum energy not gravitate as efficiently as we think it should.

Primordial Fluctuations

The ΛCDM description of cosmology is very successful, but only if the universe is started off in a very particular and unusual initial state. This is because the CMB is seen to have an almost uniform temperature in all directions in the sky, even though in the standard cosmology there has not yet been enough time in the universe's history to have this temperature equilibrate and so everywhere be the same.

Furthermore, small variations in the temperature have also been measured (at a level of one part in 10^5) and these show correlations across the sky over regions that are too large to have a causal explanation in the time available so far since the initial Big Bang.

In the Standard Model both of these would have to be accepted as a very unusual, contrived and unexplained initial condition for the initial universe. More likely the extrapolations into the past are incorrect because they miss some sort of new physics, which dynamically can explain the Hot Big Bang's peculiar initial conditions.

Baryogenesis

Another peculiar initial condition that ΛCDM cosmology requires is that the initial universe must for some reason have an incredibly tiny (but nonzero) excess of baryons over antibaryons. This is because baryon number is conserved (at least to a very good approximation, given the stability of the proton), and the Universe now seems to contain protons and neutrons but very few of their antiparticles. The observed number of baryons (protons and neutrons) minus antibaryons turns out to evolve in an expanding Universe the same way the number of photons does, and the present rarity of baryons relative to CMB photons implies $\eta_b = (n_B - \overline{n}_B)/n_\gamma \sim 10^{-10}$.

Because this ratio is time-independent it was also small in the remote past when everything was much hotter. But for high enough temperatures the baryons are relativistic and so n_B and $\overline{n}_B$ are both comparable to n_γ. This means that there must initially have been an

extremely tiny difference in the abundance of baryons and antibaryons, in order to have the earlier antibaryons annihilate with baryons as the Universe cooled, while still leaving just a few baryons left over. Why this should have happened is a puzzle.

If baryon number eventually proves not to be exactly conserved then it may eventually be possible to understand how a world with initially zero baryon number might dynamically evolve to end up with a net baryon number, and although no compelling picture for this yet exists any such an understanding certainly requires ingredients that go beyond the Standard Model.

At this writing it is not yet known how any of these five problems may ultimately be resolved, nor how the Standard Model's flavour and generation puzzles will eventually be understood. But because these represent the very few places where there is real evidence that the Standard Model's predictions fail they provide the starting points for most explorations of what might lie beyond.

Units

This appendix collects together a list of useful conversions between conventional units and fundamental units (with energies measured in eV).

1. Length and Time

$1/M_p\ (= G/\hbar c)^{\frac{1}{2}}$	=	8.1897×10^{-29}	c^2/eV	=	1.6161×10^{-35}	mc/$\hbar$
$1/m_p$	=	1.0658×10^{-9}	c^2/eV	=	2.1031×10^{-16}	mc/$\hbar$
1 fm	=	5.06773×10^{-9}	$\hbar c$/eV	=	10^{-15}	m
$1/m_e$	=	1.957×10^{-6}	c^2/eV	=	3.8616×10^{-13}	mc/$\hbar$
$a_0\ (= 1/\alpha m_e)$	=	2.6818×10^{-4}	c^2/eV	=	5.2918×10^{-11}	mc/$\hbar$
1 A	=	5.06773×10^{-4}	$\hbar c$/eV	=	10^{-10}	m
1 nm	=	5.06773×10^{-3}	$\hbar c$/eV	=	10^{-9}	m
1 μm	=	5.06773	$\hbar c$/eV	=	10^{-6}	m
1 cm	=	5.06773×10^{4}	$\hbar c$/eV	=	0.01	m
1 m	=	5.06773×10^{6}	$\hbar c$/eV	=	1	m
1 km	=	5.06773×10^{9}	$\hbar c$/eV	=	10^{3}	m
1 sec	=	1.51927×10^{15}	$\hbar$/eV	=	2.99792×10^{8}	m/c
1 min	=	9.11562×10^{16}	$\hbar$/eV	=	1.79875×10^{10}	m/c
1 hr	=	5.46937×10^{18}	$\hbar$/eV	=	1.07925×10^{12}	m/c
1 day	=	1.31265×10^{20}	$\hbar$/eV	=	2.59020×10^{13}	m/c
1 yr	=	4.795×10^{22}	$\hbar$/eV	=	9.461×10^{15}	m/c
1 pc	=	1.564×10^{23}	$\hbar c$/eV	=	3.08568×10^{16}	m
1 kpc	=	1.564×10^{26}	$\hbar c$/eV	=	3.08568×10^{19}	m
1 Mpc	=	1.564×10^{29}	$\hbar c$/eV	=	3.08568×10^{22}	m

2. Microscopic Energy and Mass

1 eV	=	10^{-9}	GeV	=	5.06773×10^{6}	$\hbar c/\mathrm{m}$
1 keV	=	10^{-6}	GeV	=	5.06773×10^{9}	$\hbar c/\mathrm{m}$
1 MeV	=	10^{-3}	GeV	=	5.06773×10^{12}	$\hbar c/\mathrm{m}$
1 GeV	=	1	GeV	=	5.06773×10^{15}	$\hbar c/\mathrm{m}$
αm_e	=	3.7289×10^{-6}	GeV/c^2	=	1.8897×10^{10}	$\hbar/\mathrm{m}c$
m_e	=	5.10999×10^{-4}	GeV/c^2	=	2.5896×10^{12}	$\hbar/\mathrm{m}c$
	=	9.10939×10^{-28}	g			
m_p	=	0.938272	GeV/c^2	=	4.75491×10^{15}	$\hbar/\mathrm{m}c$
	=	1.67262×10^{-24}	g			
	=	1.83615×10^{3}	m_e			
$M_p = (\hbar c/G)^{\frac{1}{2}}$	=	1.22105×10^{19}	GeV/c^2	=	6.1879×10^{34}	$\hbar/\mathrm{m}c$
	=	2.17671×10^{-5}	g			
	=	1.30138×10^{19}	m_p			
$\hat{M}_p = (\hbar c/8\pi G)^{\frac{1}{2}}$	=	2.43564×10^{18}	GeV/c^2	=	1.23431×10^{34}	$\hbar/\mathrm{m}c$
	=	4.34191×10^{-6}	g			
	=	2.59588×10^{18}	m_p			

3. Ordinary Units Expressed Microscopically

1 g	=	5.60959×10^{23}	GeV/c^2	=	2.84279×10^{39}	$\hbar/\mathrm{m}c$
1 kg	=	5.60959×10^{26}	GeV/c^2	=	2.84279×10^{42}	$\hbar/\mathrm{m}c$
1 Joule = 1 kg $\mathrm{m}^2/\mathrm{s}^2$	=	6.24151×10^{9}	GeV	=	3.16303×10^{25}	$\hbar c/\mathrm{m}$
1 erg = 1 g $\mathrm{cm}^2/\mathrm{s}^2$	=	6.24151×10^{2}	GeV	=	3.16303×10^{18}	$\hbar c/\mathrm{m}$
$= 10^{-7}$ J						
1 Newton = 1 kg m/s^2	=	1.23162×10^{-6}	$\mathrm{GeV}^2/\hbar c$	=	3.16303×10^{25}	$\hbar c/\mathrm{m}^2$
	=	1.23162×10^{12}	$\mathrm{eV}^2/\hbar c$			
1 dyne = 1 g cm/s^2	=	1.23162×10^{-11}	$\mathrm{GeV}^2/\hbar c$	=	3.16303×10^{20}	$\hbar c/\mathrm{m}^2$
$= 10^{-5}$ N	=	1.23162×10^{7}	$\mathrm{eV}^2/\hbar c$			
1 Watt = 1 J/s	=	4.10824×10^{-15}	$\mathrm{GeV}^2/\hbar$	=	1.05507×10^{17}	$\hbar c^2/\mathrm{m}^2$
	=	4.10824×10^{3}	$\mathrm{eV}^2/\hbar$			
1 Hz = 1/s	=	6.5821×10^{-25}	$\mathrm{GeV}/\hbar$	=	3.3356×10^{-9}	c/m
1 Kelvin	=	8.61742×10^{-14}	GeV/k_B	=	4.36707×10^{2}	$\hbar c/\mathrm{m}k_B$
	=	8.61742×10^{-5}	eV/k_B	=	$1/11604.4$	eV/k_B

4. Electromagnetic Units

1 Coulomb	=	6.24151×10^{18}	e			
1 Volt = 1 J/C	=	1	eV/e	=	5.06773×10^{6}	$\hbar c/\mathrm{m}e$
	=	10^{-9}	GeV/e			
1 Farad = 1 C/V	=	6.24151×10^{18}	e^2/eV	=	1.23162×10^{12}	$\mathrm{m}e^2/\hbar c$
1 Ampere = 1 C/s	=	4.10824×10^{3}	$\mathrm{eV}e/\hbar$	=	2.08194×10^{10}	ec/m
1 Ohm = 1 V/A	=	2.43413×10^{-4}	$\hbar/e^2$			
1 Mho = 1/Ohm	=	4.10824×10^{3}	$e^2/\hbar$			
1 Weber = 1 V s	=	1.51927×10^{15}	$\hbar/e$			
1 Tesla = 1 Weber/m^2	=	59.1572	$\mathrm{eV}^2/\hbar ec^2$	=	1.51927×10^{15}	$\hbar/e\mathrm{m}^2$
1 Gauss = 10^{-4} Tesla	=	5.91572×10^{-3}	$\mathrm{eV}^2/\hbar ec^2$	=	1.51927×10^{11}	$\hbar/e\mathrm{m}^2$
$\phi_0 = 2\pi\hbar/e$	=	6.28319	$\hbar/e$	=	4.13567×10^{-15}	Weber
				=	$1/(2.418 \times 10^{14}\ )$	Weber
$\epsilon_0 = 8.854 \times 10^{-12}$ F/m	=	10.905	$e^2/\hbar c$			
$\mu_0 = 4\pi \times 10^{-7}$ N/A^2	=	0.0917012	$\hbar/ce^2$		$\epsilon_0\mu_0 = 1/c^2$	
$\alpha = e^2/(4\pi\epsilon_0\hbar c)$	=	7.2974×10^{-3}			$1/\alpha = 137.036$	

Here is a selection of undergraduate textbooks on subatomic physics.

1. David Griffiths, *Introduction to Elementary Particles*, Wiley-VCH, 2010.
2. Ernest Henley and Alejandro Garcia, *Subatomic Physics*, Wiley-VCH, 2010.

Other useful references (to be completed...)

References

[1] C.P. Burgess and G.D. Moore, *The Standard Model: A Primer*, Cambridge University Press.