Vol. 5 – MODERN PHYSICS

The Undergraduate Companion to Theoretical Physics

Andrea Kouta Dagnino[‡]

[‡]*Open University, Milton Keynes, UK.*

E-mail: k.y.dagnino@gmail.com

Contents

The enchanting charms of this sublime science reveal only to those who have the courage to go deeply into it.

— Carl Friedrich Gauss

References

Several textbooks, online courses/resources were referenced heavily (to the extend of making this text completely unoriginal, yet hopefully helpful for revision) throughout the writing of these lecture notes. Using a typical bibliography (research paper style) would be a formidable task.Pinpointing exactly where each reference as been used is quite difficult for such a large and well-referenced subject, and would probably change the writing style to a far too formal one for lecture notes. Therefore we instead list the most relevant below giving a brief comment on which topics they were mostly used for:

- A. Steane *Relativity made Relatively Easy*

  Fantastic for a first introduction to special and general relativity with mathematical rigor (4-vector approach). Virtually the only reference needed.

- N. Woodhouse *Special Relativity*

  Another great, succinct introduction to Special relativity.

Part I

Relativity

Basic postulates of Special relativity

1

1.1 Reference frames

What is a frame of reference?

Consider an scaffolding of ruler sticks arranged in space in such a way as to denote every point in space with a set of coordinates (x, y, z), and endowed with a clock keeping track of time (by some physical, periodic phenomenon, such as a fixed number of radiative transitions in a caesium-133 atom).

Such an object is known as a frame of reference, with each space-time point (t, x, y, z), known as **events**, being specified. An inertial frame of reference where an object which is not acted upon by an external force moves at a constant velocity. In other words, it is a frame where Newton's first law holds (thus ruling out accelerating frames of references where fictitious forces are not considered to be external forces).

In classical physics, inertial frames of references satisfy galilean transformations. Consider two frames $\mathcal{S}$ and $\mathcal{S}'$ with coordinates (t, x, y, z) and (t', x', y', z'), with $\mathcal{S}'$ moving with velocity $\mathbf{v} = v_x\mathbf{x} + v_y\mathbf{y} + v_z\mathbf{z}$ as measured in $\mathcal{S}$. Then, the following transformation law is satisfied in galilean relativity:

$$\begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -v_x & 1 & 0 & 0 \\ -v_y & 0 & 1 & 0 \\ -v_z & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} \tag{1.1.1}$$

It is paramount to note that the time parameter is not affected at all by this transformation, in classical physics all clocks are assumed to be synchronized, even if they are moving relative to each other.

Maxwell vs Newton

This however leads to several contradictions and paradoxical conclusions, especially when put to the test with Maxwell's electromagnetism. For example, consider an electromagnetic wave $\mathbf{E} = \mathbf{E}_0 \sin(\mathbf{k} \cdot \mathbf{x} - \omega t)$ travelling at c as measured in the inertial frame $\mathcal{S}$. In the frame $\mathcal{S}'$, the same wave will be of the form $\mathbf{E}' = \mathbf{E}'_0 \sin(\mathbf{k}' \cdot \mathbf{x}' - \omega' t')$. We now argue that the phase of a plane wave must be an invariant quantity under a change of frame, since everyone must agree on how many crests a wave has undergone in a certain time/distance within their own frame. Consequently, we need

$$\mathbf{k}' \cdot \mathbf{x}' - \omega' t' = \mathbf{k}' \cdot \mathbf{x}' - \omega' t' \tag{1.1.2}$$

$$= \mathbf{k}' \cdot \mathbf{x} - (\mathbf{k}' \cdot \mathbf{v} + \omega')t \tag{1.1.3}$$

from which we identify $\mathbf{k}' = \mathbf{k}$ and $\omega = \mathbf{k}' \cdot \mathbf{v} + \omega' t'$. As we let $v \to c$, the observer in $\mathcal{S}'$ will observe a frozen wave with no time-dependence. This clearly isn't a plane wave solution to Maxwell's equations. So are we to believe that Maxwell's equations are only true in a specific frame of reference, the so-called aether?

The Aether

We define the aether as the frame of reference (if it even exists) in which light propagates at the conventional speed of light $c \approx 3 \times 10^8$ m/s.

Consider the following experiment. A person and a mirror are placed on the ends of a platform of length L moving at a speed $v_p \ll c$ relative to the aether. The platform is oriented so that when at rest (relative to the aether), a light beam travelling between its end has speed c. The observer sends a light beam to the mirror, which reflects back and is detected after some time. If the platform is moving along the distance between the observer and the mirror, then this time interval will be:

$$t_1 = \frac{L}{c + v_p} + \frac{L}{c - v_p} \approx \frac{2L}{c}\left(1 + \frac{v_p^2}{c^2}\right) \tag{1.1.4}$$

while if the platform is moving perpendicular to the distance L, then:

$$t_2 = \frac{2L}{\sqrt{c^2 - v_p^2}} \approx \frac{2L}{c}\left(1 + \frac{v_p^2}{2c^2}\right) \tag{1.1.5}$$

There will be a noticeable difference between these time intervals:

$$\Delta t = t_1 - t_2 \approx \frac{L v_p^2}{c^3} \tag{1.1.6}$$

which would cause a beam travelling in the parallel direction to interfere with a beam travelling in the perpendicular direction.

In the Michelson interferometer, a beam splitter is used to split a beam into two travelling in perpendicular directions, and which will interfere according to our above argument when recombining. However, no such interference effects were ever observed.

To explain this shortcoming of Galilean relativity, Lorentz and Fitzgerald argued that the aether could exert some sort of pressure on objects travelling within it, causing a contraction in its direction of motion by a factor γ:

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}} \implies L \to \gamma L \tag{1.1.7}$$

$$\implies t_2 = \frac{2L/c}{1 - v^2/c^2} = t_1 \tag{1.1.8}$$

Such an explanation, although numerically correct, fails to give the proper picture as to why such a contraction should occur. The correct explanation would ultimately arrive with Einstein.

1.2 Fundamental postulates and definitions

Postulates

The basic postulates of special relativity are the following:

(i) **Principle of relativity**: all inertial frames of reference are equivalent, and the laws of physics apply equally.

(ii) **Light speed**: the speed of light in vacuum is c irrespective of its source.

The first postulate is shared with Newtonian physics. A nice way to put it is "if you can juggle at rest, you can also juggle in an IRF", or alternatively "a blind man cannot tell if they are moving in an IRF". The second postulate, on the other hand, is shared with electromagnetism.

The problem of synchronization

We now tackle the question of synchronizing clocks. Suppose an observer sends a light beam at time t_1. It gets reflected by a mirror at an event A and reaches the observer at some time t_2. How do we synchronize the mirror's clock with the observer's clock? If we assume that light travels equally in all directions in vacuum (i.e. space is isotropic) then we can claim that the light beam reached the mirror at $\tau = \frac{1}{2}(t_1 + t_2)$ thus travelling a distance $c\tau = \frac{1}{2}c(t_1 + t_2)$.

Note however that this is just a convention. There is no way to measure the one-way speed of light and hence no way to know exactly when the light beam hit the mirror. Luckily for special relativity, it makes no difference whether or not the one way speed of light is c or some other value. Suppose that for some reason light travels at $c/2$ in the AB direction and instantaneously in the BA direction. An observer is placed at A, and another at B. Their clocks may or may not be synchronized.

At $t_0^A = 0$, the observer at A sends a message to B asking "what does your clock read". The observer at B will receive this message at $t_1^A = \frac{2l}{c}$ in A's clock, and some t_1^B in B's clock. B can respond and instantly and say "t_1^B", which will arrive at $t_2^A = t_1^A$. The observer at A then erroneously changes his clock to $t_3^A = t_1^B + \frac{l}{c}$, thinking that the message must have taken $\frac{l}{c}$ seconds to arrive since it was sent by B. He sends a message saying that his clock now reads $t_1^B + \frac{l}{c}$, arriving at $t_1^B + \frac{2l}{c}$. B then thinks that this makes sense, for A's message must have taken $\frac{l}{c}$ second to arrive.

As can be seen, even though their messages were travelling at different speeds, there were no contradictions in assuming that the one-way speed of light was c. With this convention in mind, then two people can synchronize their clocks by sending a light beam to another observer sitting exactly midway between them.

1.3 Space-time diagrams

An extremely useful tool in special relativity are space-time diagrams. It is common convention to place ct on the z-axis and x, y on the x, y-axes. A trajectory in this space is known as a **worldline**. We can revisit the problem of synchronization using these space-time diagrams. Consider two frames $\mathcal{S}$ and $\mathcal{S}'$ moving relative to each other at speed v. Three observers, A, B, C are in the frame $\mathcal{S}'$ separated by 1 unit each, and initially set their clocks so that $t = t' = 0$. In the $\mathcal{S}'$ frame, x_A, x_B, x_C's world-lines would satisfy $x' = 0, x' = 1, x' = 2$ respectively. To synchronize their clocks according to Einstein's convention, A and C must

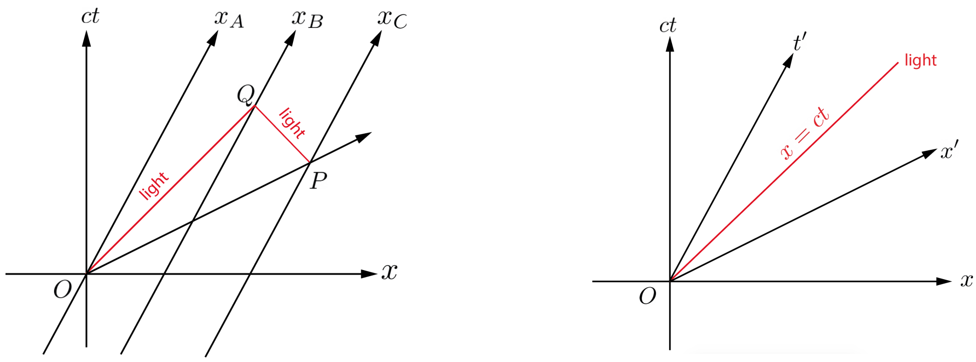


Figure 1.1. Synchronization of clocks

send a light beam to B. If their clocks are synchronized, then B will receive the signals simultaneously, making O and P synchronous in the $\mathcal{S}'$. The point P will thus also be a $t' = 0$ point since it is synchronized with O where $t' = 0$.

To find Q, we solve:

$$ct_Q = vt_Q + 1 \implies t_Q = \frac{1}{c - v} \implies x_Q = \frac{v}{c - v} + 1 \tag{1.3.1}$$

Now QP must have the form $x = c_1 - ct$ where c_1 can be found by imposing that Q lies on the line:

$$\frac{v}{c - v} + 1 = c_1 - \frac{c}{c - v} \implies c_1 = \frac{2c}{c - v} \tag{1.3.2}$$

so that P has coordinates satisfying:

$$\frac{2c}{c - v} - ct_P = vt_P + 2 \implies t_P = \frac{2v}{c^2 - v^2} \implies x_P = \frac{2c^2}{c^2 - v^2} \tag{1.3.3}$$

Consequently, the line OP for which $t' = 0$ must satisfy:

$$ct = \frac{v}{c}x \iff x = \frac{c}{v}ct \tag{1.3.4}$$

We may therefore label the line OP as the x' axis. In the $\mathcal{S}'$ frame we therefore have two tilted axes, which are reflections of each other along $x = ct$.

1.4 Fundamental consequences

Loss of simultaneity

Consider a light bulb on a moving. Observer B is inside the train while observer A is outside, they are moving at a speed v relative to each other. Two receivers are on either side of the light bulb at a distance l, and will activate when hit by a light ray.

In B's frame, the two receivers will clearly activate simultaneously after time $t_1 = t_2 = \frac{l}{c}$. In A's frame, the light from the bulb travels at speed c, but the receivers are also moving with speed v to the right. Consequently, receiver 1 will activate first after time $t_1 = \frac{l}{c+v}$ while the second will activate after time $t_2 = \frac{l}{c-v}$. The two events are not simultaneous for A even though they are for B.

This is a clear example of simultaneity being broken for two inertial observers.

We can view this in the form of a space-time diagram:

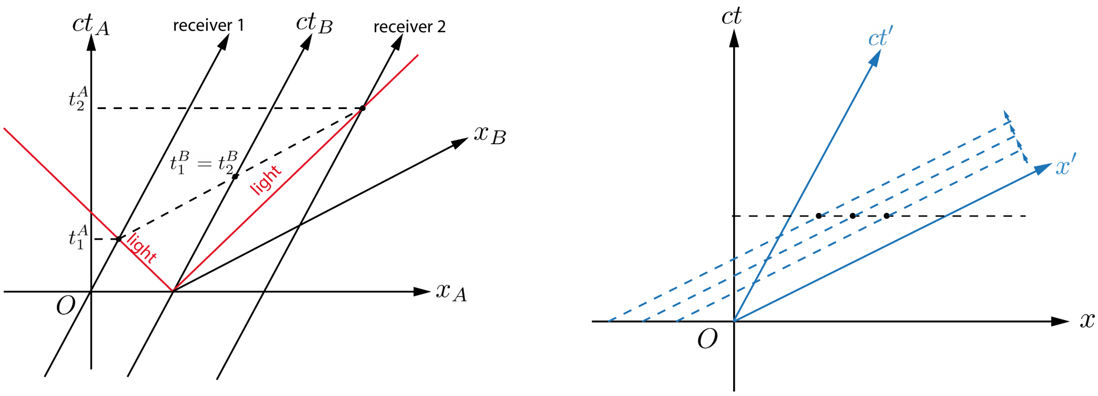


Figure 1.2. Frame dependence of simultaneity

One can also view the loss of simultaneity as a result of the "moving" observer's x'-axis being tilted. Indeed, if we envision a line parallel to the x'-axis moving along the ct'-axis, then clearly three events that are simultaneous in the stationary frame will be crossed at different times in the moving frame.

Time dilation

Consider once again a train containing an observer A moving to with speed v to the right relative to an observer B. The train has a mirror attached to its ceiling at a height h, and the observers have synchronized their clocks at time $t = 0$.

Observer A sends a light beam to the mirror at $t = 0$, in its frame it will see the reflection of the beam at time $t_A = \frac{2h}{c}$.

From observer B's point of view, the light beam has speed c along a diagonal direction, its vertical component will therefore be $\sqrt{c^2 - v^2}$. Consequently, the reflection will be ob-

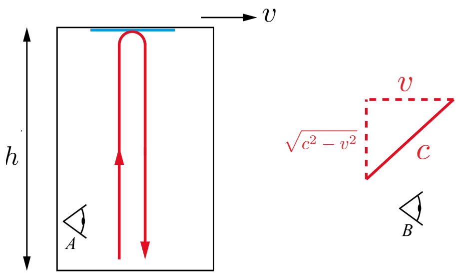


Figure 1.3. Time dilation as a result of loss of simultaneity

served at time $t_B = \frac{2h}{\sqrt{c^2-v^2}}$. Hence:

$$\boxed{t_B = \frac{t_A}{\sqrt{1-v^2/c^2}}} \tag{1.4.1}$$

Interestingly, these two times are different, the "moving observer"'s clock will run slowly compared to the "stationary observer".

We can view this more intuitively by looking at the following comic by Tatsu Takeuchi https://www1.phys.vt.edu/~takeuchi/relativity/notes/section12.html:

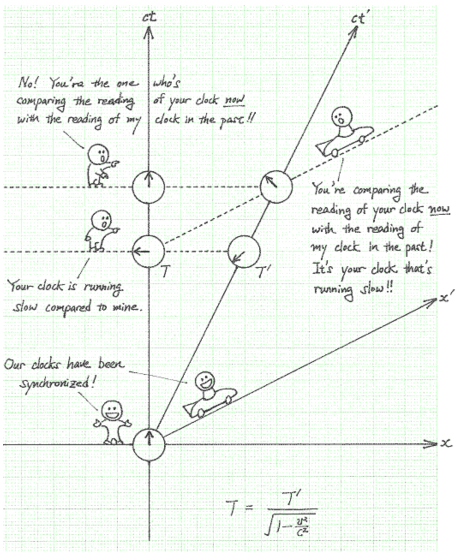


Figure 1.4. Time dilation as a result of loss of simultaneity

Due to the loss of simultaneity between two inertial observers, when they compare their clocks their definitions of simultaneity will cause them to compare their clocks with the other's clock in the past. Hence, the moving observer will always have a clock running more slowly since by the definition of simultaneity the stationary observer is looking at the moving observer's clock in the past.

Length contraction

Observer A stands on one end of a train which they have measured to have length l_A, and sends a light beam to a mirror on the other side. To them the time taken by the light beam is:

$$t_A = \frac{2l_A}{c} \tag{1.4.2}$$

For an observer B on the platform moving with speed v relative to the train, the train has length l_B, and the time taken is:

$$t_B = \frac{l_B}{c-v} + \frac{l_B}{c+v} = \frac{2l_B c}{c^2 - v^2} \tag{1.4.3}$$

since on the first trip of the light beam, the train is trying to move away from it, while on the return trip the train is moving towards it, as shown below: Consequently, using the

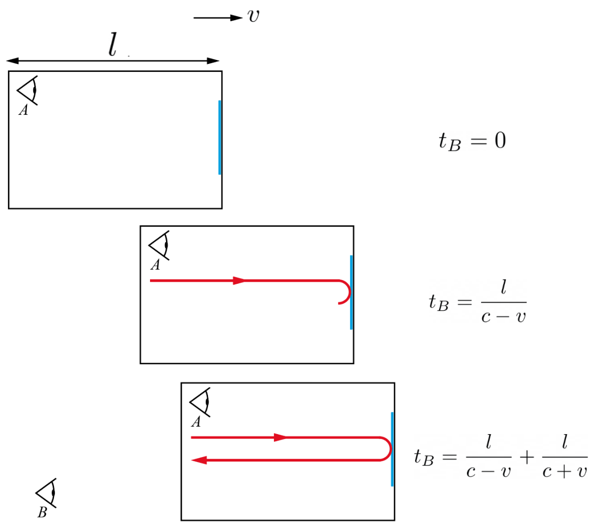


Figure 1.5. Length contraction

time dilation formula we found earlier:

$$t_B = \frac{t_A}{\sqrt{c^2 - v^2}} \implies \boxed{l_B = l_A\sqrt{1 - v^2/c^2}} \tag{1.4.4}$$

Let's consider a rod moving at speed v relative to a frame $\mathcal{S}$. We can express the position of the rod by drawing the world-lines of the front and back end of the rods, as shown below: We center the axes so that the back world-line has equation $x = vt$, while the front world-line has equation $x = vt + l$. In the still frame, the length of the rod is given by the difference in positions of the back and front world-lines at a given time t, which is $QS = l$.

In the moving frame, the length of the rod l' is given by the difference in positions of the back and front world-lines at a given time t'. From the diagram it is clear that this length

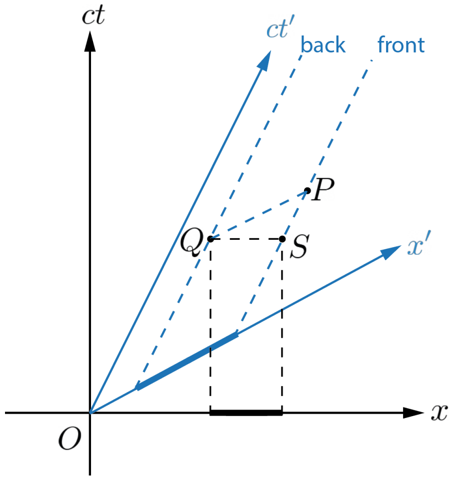


Figure 1.6. Length contraction

is shorter. Indeed:

$$x = \frac{c^2}{v} t = vt + l \implies t = \frac{v}{c^2 - v^2} l \implies x = \frac{c^2}{c^2 - v^2} l \tag{1.4.5}$$

giving a length of:

$$l' = \sqrt{c^2 t^2 - x^2} = \frac{l}{\sqrt{1 - v^2/c^2}} \tag{1.4.6}$$

The physical explanation of the minus sign will come later when we encounter the Minkowski metric, but for now let us take it as a postulate.

Interestingly, these two lengths are different, the "moving observer"'s rod will be shorter compared to the "stationary observer".

Lorentz transformations

2.1 Derivation

We now seek to find a transformation between two inertial frames $\mathcal{S} : \mathbf{x} = (ct, x, y, z)^T$ and $\mathcal{S} : \mathbf{x}' = (ct', x', y', z')^T$, where $\mathcal{S}'$ moves with velocity $\mathbf{v} = v\hat{\mathbf{e}}_x$ relative to $\mathcal{S}$. We assume that the clocks of these two frames have been synchronized at $t = t' = 0$. Firstly, by the principle of relativity if an object moves with constant velocity in one frame it must move with constant velocity in the other as well. Consequently, the transformation must be a linear one, mapping lines to lines, and keeping the origin fixed. Hence:

$$\mathbf{x}' = \Lambda\mathbf{x},\ \Lambda = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 & 0 \\ \alpha_3 & \alpha_4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.1.1}$$

where the y, z variables are left unchanged from this change of basis. Indeed, if we did have transverse effects, then this would lead to contradictions. For example, if we consider two metal pipes of equal rest diameters D_0 moving towards each other. In pipe 1's frame, pipe 2 has diameter D_2, while of course $D_1 = D_0$ is pipe 1's diameter. If $D_2 > D_0 = D_1$ (transverse length dilation), then this would mean that pipe 1 is inside pipe 2. However from pipe 2's point of view, $D_1 > D_0 = D_2$ so that pipe 2 is inside pipe 1. This is clearly a contradiction. By similar arguments, transverse length contraction is also not feasible, showing that $D_1 = D_2 = D_0$ as desired.

Now the line $x = vt$ must get mapped to $x' = 0$ so that:

$$0 = \alpha_3 ct + \alpha_4 vt \implies \alpha_3 = -\alpha_4 \frac{v}{c} \tag{2.1.2}$$

Similarly, the line $x = 0$ must get mapped to $x' = -vt'$ so that:

$$\begin{cases} -vt' = -\alpha_4 vt \\ t' = \alpha_1 t \end{cases} \implies \alpha_4 = \alpha_1 \tag{2.1.3}$$

Also, by the Light speed postulate, $x = ct$ gets mapped to $x' = ct'$ so that:

$$\begin{cases} x' = ct' = -\alpha_4 vt + \alpha_4 ct \\ ct' = \alpha_4 ct + \alpha_2 ct \end{cases} \implies \alpha_2 = -\alpha_4 \frac{v}{c} = \alpha_3 \tag{2.1.4}$$

Consequently:

$$\Lambda = \alpha_4 \begin{pmatrix} 1 & -\frac{v}{c} & 0 & 0 \\ -\frac{v}{c} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.1.5}$$

Finally, we use the principle of relativity. We know that from the perspective of $\mathcal{S}'$, it is $\mathcal{S}$ that moves with velocity $\mathbf{v} = -v\hat{\mathbf{e}_x}$. Consequently, since $\mathbf{x} = \Lambda^{-1}\mathbf{x}'$, we should have that $\Lambda(v) = \Lambda^{-1}(v)$, and thus:

$$\Lambda^{-1} = \frac{1}{\alpha_4\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & \frac{v}{c} & 0 & 0 \\ \frac{v}{c} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \alpha_4 \begin{pmatrix} 1 & \frac{v}{c} & 0 & 0 \\ \frac{v}{c} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.1.6}$$

$$\iff \alpha_4 = \frac{1}{\sqrt{1 - v^2/c^2}} \equiv \gamma(v) \tag{2.1.7}$$

Consequently, the transformation from $\mathcal{S}$ to $\mathcal{S}'$, known as a **Lorentz transformation**, can be written as:

$$\boxed{\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma(v) & -\gamma(v)\frac{v}{c} & 0 & 0 \\ -\gamma(v)\frac{v}{c} & \gamma(v) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}} \tag{2.1.8}$$

or alternatively:

$$t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{2.1.9}$$

$$x' = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{2.1.10}$$

$$y' = y \tag{2.1.11}$$

$$z' = z \tag{2.1.12}$$

In three dimensions it is easy to see how they generalize to:

$$t' = \gamma_v\left(t - \frac{\mathbf{r}\cdot\mathbf{v}}{c^2}\right) \tag{2.1.13}$$

$$\mathbf{r}'_{\parallel} = \gamma_v\left(\mathbf{r}_{\parallel} - \mathbf{v}t\right) \tag{2.1.14}$$

$$\mathbf{r}'_{\perp} = \mathbf{r}_{\perp} \tag{2.1.15}$$

2.2 Velocity addition

We know that when velocities are measured in the same frame, they add in the typical Galilean way. However, how do we deal with velocities being measured in different frames?

Longitudinal addition

Suppose we have a frame $\mathcal{S}$ in which an observer A measures another frame $\mathcal{S}'$ moving at speed v to the right. Another observer B is inside $\mathcal{S}'$ and measures the speed of a ball moving to the right to be u. What will the speed w of the ball be in $\mathcal{S}$?

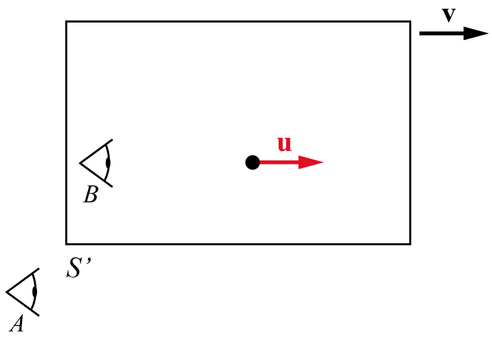


Figure 2.1. Velocity addition

We have that if the ball follows a wordline $(ct, x, 0)$ in frame $\mathcal{S}$ and $(ct', x', 0)$ in $\mathcal{S}'$, then:

$$w = \frac{x}{t} = \frac{x' + vt'}{t' + \frac{v}{c^2}x'} = \frac{u + v}{1 + \frac{uv}{c^2}} \tag{2.2.1}$$

Transverse addition

Suppose now that the ball moves in the transversally in $\mathcal{S}'$.

If the ball follows a wordline $(ct', u_x t', u_y t', u_z t')$ in $\mathcal{S}'$ then in $\mathcal{S}$ it follows a wordline (ct, x, y, z) where:

$$t = \gamma(v)(t' + \frac{u_x v}{c^2}t') \tag{2.2.2}$$
$$x = \gamma(v)(u_x t' + vt') \tag{2.2.3}$$
$$y = u_y t' \tag{2.2.4}$$
$$z = u_z t' \tag{2.2.5}$$

Consequently:

$$w_x = \frac{u_x + v}{1 + \frac{u_x v}{c^2}} \tag{2.2.6}$$
$$w_y = \frac{u_y}{\gamma(v)(1 + \frac{u_x v}{c^2})} \tag{2.2.7}$$
$$w_z = \frac{u_z}{\gamma(v)(1 + \frac{u_x v}{c^2})} \tag{2.2.8}$$

More generally, for a frame $\mathcal{S}'$ moving with velocity $\mathbf{v}$ relative to $\mathcal{S}$, if the ball moves with velocity $\mathbf{u}$ in $\mathcal{S}'$ then $\mathcal{S}$ measures:

$$\mathbf{w}_{\parallel} = \frac{\mathbf{u}_{\parallel} + \mathbf{v}}{1 + \frac{\mathbf{u}\cdot\mathbf{v}}{c^2}}, \qquad \mathbf{w}_{\perp} = \frac{\mathbf{u}_{\perp}}{\gamma(v)(1 + \frac{\mathbf{u}\cdot\mathbf{v}}{c^2})} \tag{2.2.9}$$

Rapidity

Another way to derive this result is using a quantity known as the rapidity ρ satisfying $\cosh\rho = \gamma$, $\sinh\rho = \gamma\frac{v}{c}$. The Lorentz transformation can now be written in a handy way:

$$\Lambda(\rho) = \begin{pmatrix} \cosh\rho & -\sinh\rho & 0 & 0 \\ -\sinh\rho & \cosh\rho & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.2.10}$$

Due to the additivity of $\cosh$ and $\sinh$, the composition of Lorentz transformations is simplified. Suppose in a frame $\mathcal{S}$ we measure a rapidity ρ_1 for frame $\mathcal{S}'$ in which the ball has rapidity ρ_2. Then:

$$\Lambda(\rho_2)\Lambda(\rho_1) = \begin{pmatrix} \cosh\rho_2 & -\sinh\rho_2 & 0 & 0 \\ -\sinh\rho_2 & \cosh\rho_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cosh\rho_1 & -\sinh\rho_1 & 0 & 0 \\ -\sinh\rho_1 & \cosh\rho_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.2.11}$$

$$= \begin{pmatrix} \cosh(\rho_1+\rho_2) & -\sinh(\rho_1+\rho_2) & 0 & 0 \\ -\sinh(\rho_1+\rho_2) & \cosh(\rho_1+\rho_2) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.2.12}$$

Consequently the rapidity of the ball in the frame $\mathcal{S}$ is $\rho \equiv \rho_1 + \rho_2$ implying that:

$$\tanh\rho = \tanh(\rho_1+\rho_2) = \frac{\tanh\rho_1 + \tanh\rho_2}{1 + \tanh\rho_1 \tanh\rho_2} \tag{2.2.13}$$

and recalling that $\tanh\rho = \frac{w}{c}$, $\tanh\rho_1 = \frac{v}{c}$, $\tanh\rho_2 = \frac{u}{c}$ we finally get the velocity addition rule:

$$w = \frac{u+v}{1+\frac{uv}{c^2}} \tag{2.2.14}$$

The ease with which we can combine Lorentz transformations is once again reminiscent of how one can compose rotations in a similar fashion. In the case of typical rotations, the rapidity ρ would be substituted by the

This makes sense, since in a space-time diagram $\tanh\rho$ corresponds to $\tan\theta$ where θ is the angle between the stationary and moving frames' axes.

The use of hyperbolic trigonometric functions allows us to sum angles the way we would conventionally do in euclidean geometry, only that angles now correspond to rapidities (see chapter on spinors for more details).

Rapidities also have a physical interpretation related to classical acceleration. Consider a rocket moving at speed v relative to frame $\mathcal{S}$ and with acceleration a. At time $t + dt$ the rocket is moving with velocity adt relative to its rest frame at time t. Using velocity addition, in the frame $\mathcal{S}$ we have that:

$$v(t+dt) = \frac{v(t) + adt}{1 + v(t)adt/c^2} \approx v(t) + adt - \frac{v(t)^2}{c^2}adt \tag{2.2.15}$$

$$\implies \frac{dv(t)}{dt} = a\left(1 - \frac{v(t)^2}{c^2}\right) \tag{2.2.16}$$

$$\implies \frac{v(t)}{c} = \tanh\left(\frac{1}{c}\int_0^t adt\right) = \tanh\rho \tag{2.2.17}$$

so that:

$$\boxed{\rho = \frac{1}{c}\int_0^t adt \iff \frac{d\rho}{dt} = \frac{a}{c}} \tag{2.2.18}$$

2.3 Lorentz invariance

The quantity $\mathbf{x} = (ct, x, y, z)^T$ is known as a 4-vector, any quantity that transforms as $\mathbf{x}$ under Lorentz boosts, that is through $\mathbf{x}' = \Lambda\mathbf{x}$ is known as a 4-vector. The coordinates of a 4-vector are denoted by a greek script, typically μ or ν running from 0 to 3.

A quantity is said to be Lorentz invariant if it is left unchanged under Lorentz transformation. In Newtonian mechanics, the length of a vector with Euclidean metric is invariant under rotations. This allows us to express the laws of mechanics in a frame-independent way. In a similar way it is useful to find quantities related to 4-vectors that are frame-independent in special relativity.

As one would guess from looking at the, the typical Euclidean length of $\mathbf{x}$ vector is not invariant. Indeed:

$$\mathsf{X}^T\mathsf{X} = (ct)^2 + x^2 + y^2 + z^2 \tag{2.3.1}$$

while:

$$\mathsf{X}'^T\mathsf{X}' = (\Lambda\mathsf{X})^T(\Lambda\mathsf{X}) = \mathsf{X}^T\Lambda^T\Lambda\mathsf{X} = \mathsf{X}^T\Lambda^2\mathsf{X} \tag{2.3.2}$$

where we used the symmetry of Λ. So clearly the notion of length in Euclidean geometry will not do.

Let us impose a metric $g = [\eta_{\mu\nu}]$ such that the norm of a 4-vector in this metric is Lorentz-invariant. In other words, we need the quadratic form:

$$X_\mu X^\mu = \mathsf{X}^T g\mathsf{X} = \eta_{\mu\nu}X^\mu X^\nu \tag{2.3.3}$$

and

$$X'_\mu X'^\mu = \mathsf{X}'^T g\mathsf{X} = \mathsf{X}^T(\Lambda^T g\Lambda)\mathsf{X} = X^a\Lambda^\mu_a\eta_{\mu\nu}\Lambda^\nu_b X^b \tag{2.3.4}$$

to be equal, giving an orthogonality condition:

$$\Lambda^T g \Lambda = g \iff \eta_{ab} = \Lambda^\mu_a \eta_{\mu\nu} \Lambda^\nu_b \tag{2.3.5}$$

Matrices Λ satisfying this condition form the Lorentz group, which are discussed in detail in the Mathematical methods volume. The Lorentz group has a remarkable resemblance with the rotation group $O(3)$, which satisfies a similar orthogonality condition in Euclidean space:

$$R^T \mathbb{1} R = \mathbb{1} \iff \delta_{ab} = R^i_a \delta_{ij} R^j_b \tag{2.3.6}$$

since $\mathbb{1} = [\delta_{ij}]$ is the Euclidean metric.

Going back to the postulate of light speed, we can gain insight into the form of g by imposing that two light-like separated events in one inertial frame be so in all inertial frames. In other words, if say an event with $\mathbf{x} = (ct, x, y, z)$ is light-like separated from the origin in one frame:

$$(ct)^2 - x^2 - y^2 - z^2 = 0 \tag{2.3.7}$$

then similarly:

$$(ct')^2 - x'^2 - y'^2 - z'^2 = 0 \tag{2.3.8}$$

in any other arbitrary primed frame. One should therefore choose a metric of the form:

$$g = [\eta_{\mu\nu}] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.3.9}$$

known as the **Minkowski metric** with $(+ - --)$ signature. It is easy to verify that this metric does indeed satisfy the orthogonality condition (2.3.6).

2.4 Space-time intervals

Given two events (ct_1, x_1, y_1, z_1) and (ct_2, x_2, y_2, z_2), their space-time interval is thus defined as:

$$(\Delta s)^2 = \eta_{\mu\nu} \Delta X^\mu \Delta X^\nu = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \tag{2.4.1}$$

The sign of the space-time interval between two events can give insight into their properties:

(i) if $\Delta s > 0$ then the events are **time-like** separated, that is, a physical signal could travel between the two events. It corresponds to the region contained within the light cone. Alternatively, one can find a frame where the two events occur at the same position, but there does not exist a frame where they are simultaneous.

(ii) if $\Delta s < 0$ then the events are **space-like** separated, that is, no physical signal can travel between the two events. It corresponds to the region outside the light cone. Alternatively, one can find a frame where the two events are simultaneous, but there does not exist a frame where they occur at the same position.

(iii) if $\Delta s = 0$, then the events are **light-like** separated, that is, only a light signal can travel between the two events. It corresponds to the surface of the light cone.

As can be seen from the figure below, the surfaces of constant space-time interval form hyperboloids.

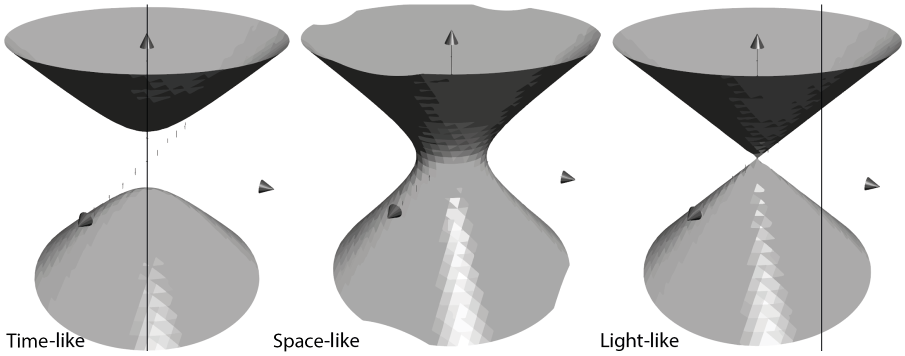


Figure 2.2. Surfaces of constant space-time interval in 2+1 space, with ct on the z-axis, and x, y in the $x - y$ plane.

Using the space-time interval, which is a Lorentz invariant quantity, we may also formally define the concepts of distance and time. For two events that are time-like separated, the distance between them is given by the proper length:

$$\Delta r = -\Delta s \tag{2.4.2}$$

Since we can find a frame $\tilde{S}$ where the events are simultaneous, we see that Δr is the distance between the events measured simultaneously in $\tilde{S}$.

For two events that are space-like separated, the time between them is given by the proper time:

$$\Delta \tau = \frac{\Delta s}{c} \tag{2.4.3}$$

2.5 4-vectors

4-velocity

Consider the world-line of a particle moving through space relative to an inertial frame. The differential proper time between any two $(ct, \mathbf{r})$ and $(c(t + dt), \mathbf{r} + d\mathbf{r})$ is:

$$\begin{aligned} d\tau = \frac{ds}{c} &= \frac{1}{c}\sqrt{g_{\mu\nu} dX^\mu dX^\nu} && (2.5.1) \\ &= \frac{1}{c}\sqrt{g_{\mu\nu}\frac{dX^\mu}{dt}\frac{dX^\nu}{dt}}dt && (2.5.2) \\ &= \frac{1}{c}\sqrt{c^2 - v^2}dt && (2.5.3) \\ &= \frac{dt}{\gamma(v)} && (2.5.4) \end{aligned}$$

where $v = \sqrt{\delta_{ij}\frac{dX^i}{dt}\frac{dX^j}{dt}}$ is the conventional 3-velocity of the particle. This allows us to find the proper time between any two events A and B on this world-line:

$$\Delta\tau = \int_A^B \frac{dt}{\gamma(v)} = \frac{\Delta t}{\gamma(v)} \tag{2.5.5}$$

as we found earlier when discussing time-dilation.

Using proper-time, we can create a 4-velocity whose norm which will be Lorentz invariant:

$$\mathsf{U} = \frac{d\mathsf{X}}{d\tau} = \frac{d}{d\tau}(ct, \mathbf{r}) = \gamma(v)\begin{pmatrix} c \\ \mathbf{v} \end{pmatrix} \tag{2.5.6}$$

Its norm is clearly:

$$||\mathsf{U}|| \equiv \mathsf{U}^T g \mathsf{U} = \gamma(v)\sqrt{c^2 - v^2} = c \tag{2.5.7}$$

which is not only Lorentz-invariant as desired, but also constant.

4-momentum

In Newtonian mechanics, momentum is defined as $\mathbf{p} = m\mathbf{v}$, where m is a Galilean-invariant quantity. Similarly, in Special relativity we can define the 4-momentum using a Lorentz-variant mass, the rest mass m_0, which is defined as the mass of the object as measured in its frame. Hence:

$$\mathsf{P} = m_0\mathbf{v} = m_0\gamma(v)\begin{pmatrix} c \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix} \tag{2.5.8}$$

where we defined:

$$E = \gamma(v)m_0c^2,\ \mathbf{p} = \gamma(v)m_0\mathbf{v} \tag{2.5.9}$$

to be the relativistic energy and momenta respectively (we shall motivate the definition for the former later).

Its norm is found to be:

$$||\mathsf{P}|| = m_0\gamma(v)\sqrt{c^2 - v^2} = m_0c \tag{2.5.10}$$

which is Lorentz invariant as desired. Consequently, we find that:

$$\boxed{E^2 - p^2c^2 = m^2c^4} \tag{2.5.11}$$

4-gradient

Note that we can write the transformation law for 4-position as:

$$X'^\nu = \Lambda^\nu_\mu X^\mu = \frac{\partial X'^\nu}{\partial X^\mu}X^\mu \tag{2.5.12}$$

$$X'_\nu = \Lambda^\mu_\nu X_\mu = \frac{\partial X^\mu}{\partial X'^\nu}X_\mu \tag{2.5.13}$$

which gives us the typical definition of contravariant and covariant vectors. It then follows that:

$$\partial'_\nu \equiv \frac{\partial}{\partial X'^\nu} = \frac{\partial X^\mu}{\partial X'^\nu}\frac{\partial}{\partial X^\mu} = \Lambda^\mu_\nu \partial_\mu \tag{2.5.14}$$

$$\partial'^\nu \equiv \frac{\partial}{\partial X'_\nu} = \frac{\partial X'^\nu}{\partial X^\mu}\frac{\partial}{\partial X_\mu} = \Lambda^\nu_\mu \partial^\mu \tag{2.5.15}$$

Hence, we see that we may define a new 4-operator $\square$, known as 4-gradient, with contravariant components ∂^μ by differentiating with respect to covariant position components:

$$\partial^\mu = \left(\frac{1}{c}\frac{\partial}{\partial t}, -\nabla\right) \tag{2.5.16}$$

and with covariant components ∂_μ by differentiating with respect to contravariant position components:

$$\partial_\mu = \left(\frac{1}{c}\frac{\partial}{\partial t}, \nabla\right) \tag{2.5.17}$$

When we operate on some Lorentz scalar ϕ with the 4-gradient, we get a 4-vector since:

$$\partial'^\nu \phi = \Lambda^\nu_\mu \partial^\mu \phi \tag{2.5.18}$$

If instead we operate on a 4-vector, then:

$$\square' \cdot \mathsf{V}' = g_{\mu\nu}\partial'^\mu V'^\nu = (\Lambda^\mu_\alpha g_{\mu\nu}\Lambda^\mu_\beta)\partial^\alpha V^\beta = g_{\alpha\beta}\partial^\alpha V^\beta \tag{2.5.19}$$

so we get a Lorentz scalar. For example, $\square \cdot \mathsf{X} = 4$.

It follows that $\square = \partial^\mu\partial_\mu$ must be a scalar operator, known as the d'Alembertian operator. It is equivalent to the classical wave operator:

$$\boxed{\square^2 \equiv \partial^\mu\partial_\mu = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2} \tag{2.5.20}$$

4-wavevector

Let us assume that the phase $\phi = \mathbf{k}\cdot\mathbf{r} - \omega t$ of a plane wave be Lorentz-invariant (this should be case, since all observers should agree on how many cycles a wave has gone through). This is a well motivated choice as we will soon explain. Noting that $\phi = (\frac{\omega}{c}, \mathbf{k}) \cdot (ct, \mathbf{r})$, one would be inclined to define the following quantity:

$$\mathsf{K} = \begin{pmatrix} \frac{\omega}{c} \\ \mathbf{k} \end{pmatrix} \tag{2.5.21}$$

To see that our instincts are justified, consider the following thought experiment. Suppose an observer in some frame measures the number of wave fronts crossing a finite volume in some time interval. The number of crests will be proportional to the measured phase. Now another observer in a frame moving relate to the initial one will still record the same number of crests even though the finite volume and time intervals will be different. Hence the measured phase must be invariant.

Taking the 4-gradient of the phase we obtain a 4-vector known as the 4-wavevector:

$$\mathsf{K} = \Box\phi = \begin{pmatrix} \frac{\omega}{c} \\ \mathbf{k} \end{pmatrix} \tag{2.5.22}$$

The norm of the 4-wavevector is:

$$||\mathsf{K}|| = \frac{\omega^2}{c^2} - k^2 = \omega^2\left(\frac{1}{c^2} - \frac{1}{v_p^2}\right) \tag{2.5.23}$$

where $v_p = \frac{\omega}{k}$ is the phase-speed of a mode ω.

2.6 The Doppler effect

Suppose in frame $\mathcal{S}'$ we have a plane wave moving in the $x'y'$ plane, making an angle θ' with the x' axis, with wave-number k' and angular frequency ω'. Hence we have that:

$$\mathsf{K}' = (\frac{\omega'}{c}, k'\cos\theta', k'\sin\theta', 0) \tag{2.6.1}$$

In the stationary frame $\mathcal{S}$, we have that:

$$\begin{pmatrix} \frac{\omega}{c} \\ k\cos\theta \\ k\sin\theta \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\omega'}{c} \\ k'\cos\theta' \\ k'\sin\theta' \\ 0 \end{pmatrix} \tag{2.6.2}$$

implying that:

$$\omega = \gamma\omega'\left(1 + \frac{v}{\omega'}k'\cos\theta'\right), \qquad \tan\theta = \frac{\sin\theta'}{\gamma(\frac{v\omega'}{k'c^2} + \cos\theta')} \tag{2.6.3}$$

Defining the phase velocity in $\mathcal{S}'$ to be $v_p = \frac{\omega'}{k'}$ then these become:

$$\omega = \gamma\omega'\left(1 + \frac{v}{v_p}\cos\theta'\right) \tag{2.6.4}$$

$$\tan\theta = \frac{\sin\theta'}{\gamma(\cos\theta' + \frac{v_p v}{c^2})} \tag{2.6.5}$$

These equations define the relativistic Doppler effect. There are two special cases of the Doppler effect, the transverse effect where $\cos\theta = 0$, and the longitudinal effect where $\cos\theta' = 1$, both of which can be understood through time dilation and length contraction.

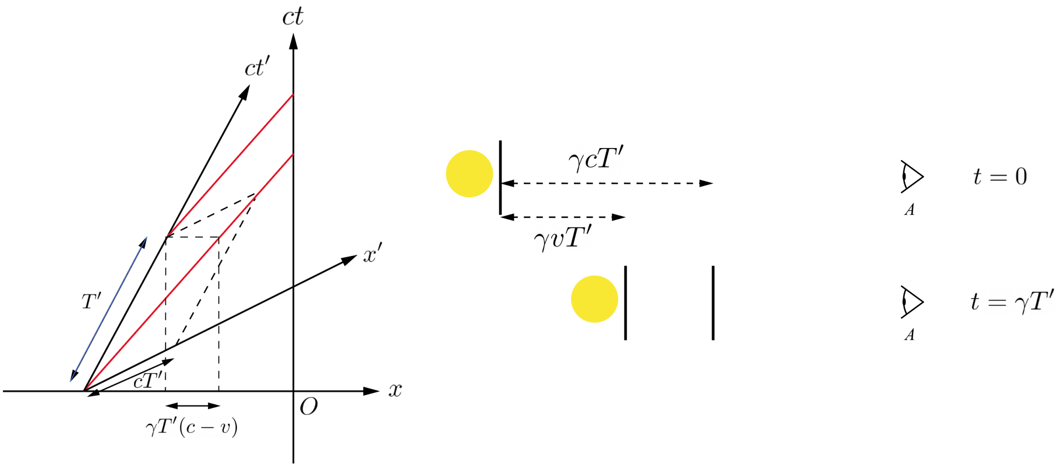


Figure 2.3. Longitudinal Doppler effect

Longitudinal Doppler effect

Here we find that:

$$\frac{\omega}{\omega'} = \sqrt{\frac{1+v/c}{1-v/c}} \tag{2.6.6}$$

We can interpret this as follows. In the source's frame $\mathcal{S}'$, the distance between two crests is $\lambda' = \frac{2\pi}{k'} = cT'$ where $T' = \frac{2\pi}{\omega'}$, so that $T = \frac{2\pi}{k'c}$. In the stationary frame $\mathcal{S}'$, we have that at time $t = 0$, a wave-front is emitted. At $t = T = \gamma T'$, then the second wave-front is emitted, but because the source is moving, the distance between the crests will be $\lambda = \gamma T'(c - v)$. Consequently:

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{\gamma c T'(1-v/c)} = \frac{k'}{\gamma T'(1-v/c)} \tag{2.6.7}$$

$$\implies \frac{\omega}{\omega'} = \frac{k}{k'} = \sqrt{\frac{1+v/c}{1-v/c}} \tag{2.6.8}$$

We can understand this through a helpful space-time diagram shown above.

Transverse Doppler effect

Here we find that $\cos\theta = 0$ and thus $\cos\theta' = -\frac{v_p v}{c^2}$. Consequently:

$$\frac{\omega}{\omega'} = \frac{1}{\gamma} \tag{2.6.9}$$

This follows clearly from applying time dilation, if the wave has period T' in $\mathcal{S}'$ then in $\mathcal{S}$ we have a period $T = \gamma T'$ and thus $\omega' = \gamma\omega \implies \frac{\omega}{\omega'} = \frac{1}{\gamma}$ as desired.

2.7 Thomas precession

Consider the following. In a frame $\mathcal{S}$ we have two squares, one moving upwards with speed u and another moving downwards with speed v. Two of their corners are labelled A and B as shown. We consider two additional frames: $\mathcal{S}$ and $\mathcal{S}''$ which are the rest frames of

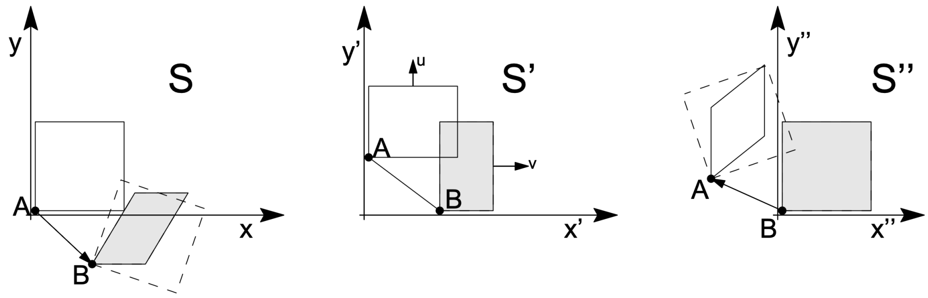


Figure 2.4. A double lorentz boost is equivalent to a single lorentz boost times a rotation. (Have to replace with my own image)

the white and gray squares respectively. We align their frames in $\mathcal{S}'$ along their respective squares.

In frame $\mathcal{S}$ velocity addition tells us that the gray square will be moving with speed $v_{\|} = u, v_{\perp} = \frac{v}{\gamma_u}$. Hence the line AB makes an angle θ with the x-axis satisfying $\tan\theta = \frac{\gamma_u u}{v}$.

Similarly, in frame $\mathcal{S}''$ velocity addition tells us that the white square will be moving with speed $u_{\|} = v, u_{\perp} = \frac{u}{\gamma_v}$. Hence the line AB makes an angle θ'' with the x-axis satisfying $\tan\theta'' = \frac{u}{\gamma_v v}$.

Clearly, these two angles are not the same. In other words, the axes of $\mathcal{S}$ and $\mathcal{S}''$ are misaligned in each other's frames but not in $\mathcal{S}'$!

We may also write that the misalignment $\Delta\theta$ satisfies:

$$\tan\Delta\theta = \frac{\frac{\gamma_u u}{v} - \frac{u}{\gamma_v v}}{1 + \frac{u}{\gamma_v v}\frac{\gamma_u u}{v}} = \frac{uv(\gamma_u\gamma_v - 1)}{\gamma_u u^2 + \gamma_v v^2} \tag{2.7.1}$$

This effect is known as Thomas precession, and the above formula applies even for non-orthogonal velocities. When we perform two successive Lorentz boosts in opposite directions, this will be equivalent to a single Lorentz boost plus an additional rotation by $\Delta\theta$.

Our rapidity statement that Lorentz boosts add up only applied because we were considering boosts in the same direction, for which $\Delta\theta = 0$.

Circular motion

Consider for example a pilot flying a plane along a circle which we model as an N sided regular polygon with internal angles $\theta = (1 - \frac{2}{N})\pi$ with N very large. At each vertex, the

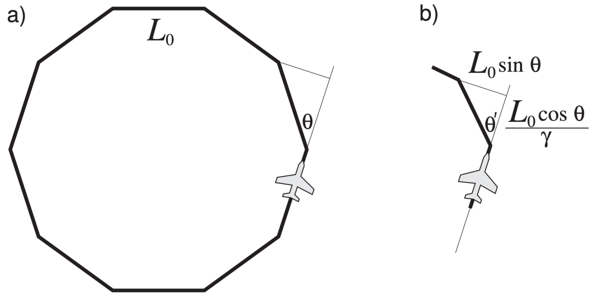


Figure 2.5. A double lorentz boost is equivalent to a single lorentz boost times a rotation. (Have to replace with my own image)

pilot must therefore rotate by an angle θ', which due to Lorentz contraction satisfies:

$$\tan\theta' = \gamma \tan\theta \implies \theta' \approx \gamma\theta \tag{2.7.2}$$

However, this means that after having gone all the way around the polygon, that is, after N rotations, the overall angle the pilot will have rotated by would be $2\pi\gamma > 2\pi$. There has been an extra rotation by $2\pi(\gamma - 1)$! This seemingly paradoxical result is of course be explained through Thomas precession.

Indeed, let us assume a momentary rest frame $\mathcal{S}'$ of the pilot. Here it is moving with velocity $\mathbf{v}$ relative to the rest frame $\mathcal{S}$ of the circle. In time $d\tau$ the pilot will be moving relative to $\mathcal{S}'$ with velocity $d\mathbf{v}_0 = \mathbf{a}_0 d\tau$ where $\mathbf{a}_0$ is the pilot's proper acceleration. Let the new instantaneous frame be $\mathcal{S}''$. It is important to note that $\mathbf{a}_0$ always points towards the center of the circle and is thus perpendicular to $\mathbf{v}_0$. Consequently, to move from time τ to

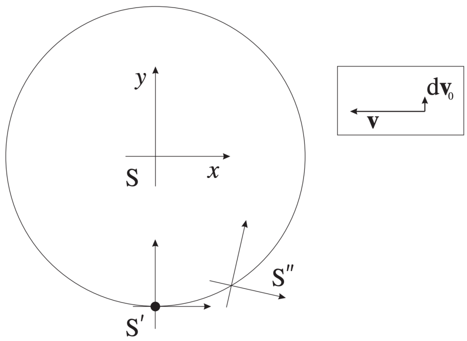


Figure 2.6. A double lorentz boost is equivalent to a single lorentz boost times a rotation. (Have to replace with my own image)

$\tau + d\tau$ we will have to perform a Lorentz boost from $\mathcal{S}$ (circle rest frame) to $\mathcal{S}'$ (pilot rest frame at τ) to $\mathcal{S}''$ (pilot rest frame at τ') along two orthogonal directions, first $\mathbf{v}_0$ and then $d\mathbf{v}$. We have already found the resulting precession angle seen from $\mathcal{S}$:

$$\tan d\theta \approx d\theta = \frac{v dv_0(\gamma_v - 1)}{\gamma_v v^2} = \left(1 - \frac{1}{\gamma_v}\right)\frac{dv_0}{v} \tag{2.7.3}$$

Finally, we substitute $dv_0 = \gamma_v dv$)by velocity addition) to find:

$$d\theta = (\gamma_v - 1)\frac{dv}{v} \implies \Delta\Theta = 2\pi(\gamma_v - 1) \tag{2.7.4}$$

as found earlier.

Tensors and the Lorentz groups

3

As was prefaced in the previous chapter, index notation is a very powerful, but sometimes quite confusing tool that is used in relativity (and most of modern physics). We have used it without giving a very thorough justification, and we should therefore reserve a chapter to discuss the intricacies of these indices, and more importantly, the objects they index, tensors. A more in-depth discussion of tensors and differential geometry is given in my Mathematical methods volume.

3.1 Vector and Dual spaces

3.2 Tensors

As an example, consider the following defining property of Lorentz matrices:

$$\Lambda^T g \Lambda = g \tag{3.2.1}$$

How do we write this in tensor notation? We have that:

$$\eta = \eta_{\alpha\beta} \boldsymbol{\varepsilon}^\alpha \otimes \boldsymbol{\varepsilon}^\beta \tag{3.2.2}$$

and:

$$\begin{aligned}
\Lambda^T g \Lambda &= (\Lambda^\mu{}_\alpha \mathbf{e}_\mu \otimes \boldsymbol{\varepsilon}^\alpha)^T (\eta_{\sigma\gamma} \boldsymbol{\varepsilon}^\sigma \otimes \boldsymbol{\varepsilon}^\gamma)(\Lambda^\nu{}_\beta \mathbf{e}_\nu \otimes \boldsymbol{\varepsilon}^\beta) && (3.2.3)\\
&= (\Lambda^\mu{}_\alpha \boldsymbol{\varepsilon}^\alpha \otimes \mathbf{e}_\mu)(\eta_{\sigma\gamma} \boldsymbol{\varepsilon}^\sigma \otimes \boldsymbol{\varepsilon}^\gamma)(\Lambda^\nu{}_\beta \mathbf{e}_\nu \otimes \boldsymbol{\varepsilon}^\beta) && (3.2.4)\\
&= \Lambda^\mu{}_\alpha \eta_{\sigma\gamma} \Lambda^\nu{}_\beta \boldsymbol{\varepsilon}^\sigma(\mathbf{e}_\mu) \boldsymbol{\varepsilon}^\gamma(\mathbf{e}_\nu) \boldsymbol{\varepsilon}^\alpha \otimes \boldsymbol{\varepsilon}^\beta && (3.2.5)\\
&= \Lambda^\mu{}_\alpha \eta_{\sigma\gamma} \Lambda^\nu{}_\beta \delta^\sigma_\mu \delta^\gamma_\nu \boldsymbol{\varepsilon}^\alpha \otimes \boldsymbol{\varepsilon}^\beta && (3.2.6)
\end{aligned}$$

implying that:

$$\boxed{\eta_{\alpha\beta} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \eta_{\mu\nu}} \tag{3.2.7}$$

which we wrote down in the previous chapter (we did not go through this very elegant reasoning, but rather argued that as $\mu, \nu = 0, 1, 2, 3$ it simply gives the correct terms). We have done this calculation in excruciating detail, but with time it should become fairly routine.

Also, let S be a $(1,1)$-tensor which we expand in the $\{\mathbf{e}_\mu\}$ and $\{\boldsymbol{\varepsilon}^\nu\}$ bases:

$$S = S^\mu{}_\nu \mathbf{e}_\mu \otimes \boldsymbol{\varepsilon}^\nu \tag{3.2.8}$$

We can take the transpose of this tensor (note that taking the transpose of a tensor only makes sense within a chosen matrix representation):

$$S^T = S^\mu{}_\nu \varepsilon^\nu \otimes \mathbf{e}_\mu = (S^T)_\nu{}^\mu \varepsilon^\nu \otimes \mathbf{e}_\mu \tag{3.2.9}$$

implying that:

$$(S^T)_\nu{}^\mu = S^\mu{}_\nu \tag{3.2.10}$$

3.3 Covariant vs. contravariant

We have found that contravariant components transform as:

$$X'^\mu = \Lambda^\mu{}_\nu X^\nu \tag{3.3.1}$$

where follow the notation in Weinberg of priming the component X, not the index. In other texts, such as Carroll, Schutz or Dirac, we prime the index:

$$X^{\mu'} = \Lambda^{\mu'}{}_\nu X^\nu \tag{3.3.2}$$

Both are perfectly fine, although it does lead to some confusion when referencing several texts! I will mostly use Weinberg's notation although whenever you see primed indices it is implicitly assumed that we are using the other convention.

We can lower the indices in (3.3.1) using the metric tensor and find that:

$$X'_\mu = \Lambda_\mu{}^\nu X_\nu \tag{3.3.3}$$

In the other notation this reads:

$$X_{\mu'} = \Lambda^\nu{}_{\mu'} X_\nu \tag{3.3.4}$$

To see why in the other notation, note that a vector itself is an abstract object and does not depend on our artificial choice of basis. Consequently:

$$\mathsf{X} = X^{\mu'} \mathbf{e}_{\mu'} = \Lambda^{\mu'}{}_\nu X^\nu \mathbf{e}_{\mu'} = X^\nu \mathbf{e}_\nu \implies \mathbf{e}_{\mu'} = \mathbf{e}_\nu (\Lambda^{-1})^\nu{}_{\mu'} \tag{3.3.5}$$

$$\implies X_{\mu'} = \left\langle X_\nu \varepsilon^\nu, \mathbf{e}_\nu (\Lambda^{-1})^\nu{}_{\mu'} \right\rangle = (\Lambda^{-1})^\nu{}_{\mu'} X_\nu = \Lambda^\nu{}_{\mu'} X_\nu \tag{3.3.6}$$

where we defined $(\Lambda^{-1})^\nu{}_{\mu'} \equiv \Lambda^\nu_{\mu'}$. This makes sense, since the inverse of a Lorentz transformation from unprimed to primed coordinates is equivalent to a Lorentz transformation from primed to unprimed coordinates. It is crucial to note that the contravariant and covariant components transform in opposite ways, their transformation matrices are inverses of each other:

$$\Lambda^\alpha{}_{\mu'} \Lambda^{\mu'}{}_\beta = \delta^\alpha_\beta \tag{3.3.7}$$

In Weinberg notation, we can derive this result using the definition of the Lorentz group:

$$\Lambda^\alpha{}_\mu \eta_{\alpha\beta} \Lambda^\beta{}_\nu = \eta_{\mu\nu} \implies \Lambda^\alpha{}_\mu \Lambda_{\alpha\nu} = \eta_{\mu\nu} \implies \Lambda^\alpha{}_\mu \Lambda_\alpha{}^\nu = \delta^\nu_\mu \tag{3.3.8}$$

implying that:

$$(\Lambda^{-1})^{\nu}{}_{\alpha} = \Lambda_{\alpha}{}^{\nu} \tag{3.3.9}$$

This, together with (3.2.10) lead to the somewhat confusing result:

$$(\Lambda^{-1})^{\mu}{}_{\alpha} = \Lambda_{\alpha}{}^{\mu} = (\Lambda^{T})^{\mu}{}_{\alpha} \tag{3.3.10}$$

This result is indeed correct, but requires some thought to be interpreted correctly. Firstly, this does not imply that $\Lambda^{-1} = \Lambda^{T}$, as this makes no sense at all (they are completely different maps). Indeed, we know that the components of Λ^{T} are $(\Lambda^{T})_{\nu}{}^{\mu}$, and consequently:

$$(\Lambda^{T})^{\mu}{}_{\alpha} = \eta^{\mu\gamma}\eta_{\sigma\alpha}(\Lambda^{T})_{\gamma}{}^{\sigma} = (\eta\Lambda^{T}\eta)^{\mu}{}_{\alpha} \tag{3.3.11}$$

so (3.3.10) becomes:

$$\Lambda^{-1} = \eta\Lambda^{T}\eta \tag{3.3.12}$$

All (3.3.10) is saying is that the Lorentz matrices are orthogonal in the Minkowski metric, which is the expression we started with in the beginning. If we instead recognize $\Lambda = [\Lambda_{\alpha}{}^{\mu}]$ then , but now the same argument must be applied to the inverse giving $\eta\Lambda^{-1}\eta = \Gamma^{T}$.

Morale of the story: you can't just equate stuff with same indices, they must have the correct index structure too!

3.4 The Lorentz group and representations

3.5 The Poincare group and representations

4

Relativistic dynamics

4.1 4-force

Transformation law

From Newton's second law, we can define the 4-force via the derivative of the 4-momentum as follows:

$$\mathsf{F} = \frac{d\mathsf{P}}{d\tau} = \left(\frac{1}{c}\frac{dE}{d\tau}, \frac{d\mathbf{p}}{d\tau}\right) \tag{4.1.1}$$

Let us define $\mathbf{f} = \frac{d\mathbf{p}}{dt}$ as the 3-force, then we find:

$$\mathsf{F} = \gamma\left(\frac{1}{c}\frac{dE}{dt}, \mathbf{f}\right) \tag{4.1.2}$$

Obviously, an invariant quantity that we can construct is:

$$\mathsf{U} \cdot \mathsf{F} = \gamma^2\left(\frac{dE}{dt} - \mathbf{u} \cdot \mathbf{f}\right) \tag{4.1.3}$$

We can calculate this quantity most easily in the particle's rest frame where $\mathbf{u} = 0$ and $E = mc^2$:

$$\mathsf{U} \cdot \mathsf{F} = \gamma^2 c^2 \frac{dm}{dt} = c^2 \frac{dm}{d\tau} \tag{4.1.4}$$

where we recast the result using invariant quantities. We see that when U and F are orthogonal, the rest mass is constant. Consequently, we get that:

$$\frac{dE}{dt} = \mathbf{u} \cdot \mathbf{f} \tag{4.1.5}$$

Such forces which go solely into changing the kinetic energy of the particle are known as **pure forces**.

Using the Lorentz transformations, it is easy to see that the 4-force transforms according to:

$$\frac{dE'}{dt'} = \frac{\frac{dE}{dt} - vf_{\|}}{1 - \mathbf{u}\cdot\mathbf{v}/c^2} \tag{4.1.6}$$

$$f'_{\|} = \frac{f_{\|} - \frac{v}{c^2}\frac{dE}{dt}}{1 - \mathbf{u}\cdot\mathbf{v}/c^2} \tag{4.1.7}$$

$$f'_{\perp} = \frac{f_{\perp}}{\gamma(v)(1 - \mathbf{u}\cdot\mathbf{v}/c^2)} \tag{4.1.8}$$

As we can see, the 3-force is not invariant at all. Now we have that for a pure 3-force **f**:

$$\mathbf{f} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(\gamma m_0 \mathbf{u}) = \gamma m_0 \mathbf{a} + m_0 \mathbf{u}\frac{d\gamma}{dt} \tag{4.1.9}$$

where $\mathbf{a} = \frac{d\mathbf{u}}{dt}$ is the usual acceleration. After some algebra one finds that:

$$\frac{d\gamma}{dt} = \frac{1}{m_0 c^2}\frac{dE}{dt} = \frac{\mathbf{u}\cdot\mathbf{f}}{m_0 c^2} \tag{4.1.10}$$

$$\mathbf{f} = \gamma m_0 \mathbf{a} + \frac{\mathbf{u}\cdot\mathbf{f}}{c^2}\mathbf{u} \tag{4.1.11}$$

giving the parallel and perpendicular components to **u**:

$$f_{\|} = \gamma m_0 a_{\|} + \frac{u^2}{c^2} f_{\|} \implies \boxed{f_{\|} = \gamma^3 m_0 a_{\|}} \tag{4.1.12}$$

and similarly:

$$\boxed{f_{\perp} = \gamma m_0 a_{\perp}} \tag{4.1.13}$$

Clearly, we see that the force acting on the particle is not necessarily parallel to its acceleration. This follows from the fact that the component $\mathbf{p}^{\perp}$ perpendicular to the force cannot change. In other words, we require:

$$p_f^{\perp} = p_i^{\perp} \implies \gamma(v_f)v_f^{\perp} = \gamma(v_i)v_i^{\perp} \tag{4.1.14}$$

so we see that the perpendicular velocity component must change as a result of the $\gamma(v)$ factor changing in the acceleration process.

The great train disaster

A train with rest length L is moving relative towards a bridge with Lorentz factor $\gamma = 3$. The bridge has a rest length of L and is divided into 3 sections of equal rest length.

From the bridge's point of view, the train gets contracted by a factor of 3 so all of the train's weight is acting on just one section, so the bridge breaks and the train falls.

The bridge's architect however states that from the train's point of view the bridge is just 100 meters long so there's no way the train could have fallen. In fact each section only had to support 1/9 the train's weight.

To resolve this paradox let's consider two frames, the rest frame of the bridge $\mathcal{S}$ and the rest frame of the train $\mathcal{S}'$. We note that the a force acting on the each train particle transforms as $f' = \gamma f$ while the weight force acting on each bridge particle transforms as $W' = W/\gamma$.

The breaking force of each section is smaller than $f = nW$ in the bridge frame, where n is the number of particles the train is made up of. The breaking force in the train frame is then smaller than $f' = \gamma nW = \gamma^2 nW'$. In other words, each section can't support $1/9$ of the train's rest weight W'.

4.2 Relativistic rockets

Consider a particle accelerating along a line. Suppose that in frame $\mathcal{S}$ the particle is moving with speed v at event A. In a proper time $d\tau$, the particle is now moving at a speed $v(t+d\tau)$ relative to $\mathcal{S}$:

$$v(t+d\tau) = \frac{v(t) + ad\tau}{1 + v(t)ad\tau/c^2} \approx v(t) + ad\tau - \frac{v(t)^2}{c^2}ad\tau \tag{4.2.1}$$

$$\implies \frac{dv(t)}{d\tau} = a\left(1 - \frac{v(t)^2}{c^2}\right) \tag{4.2.2}$$

$$\implies \frac{v(t)}{c} = \tanh\left(\frac{1}{c}\int_0^t ad\tau\right) = \tanh\rho \tag{4.2.3}$$

implying that:

$$\frac{d\rho}{d\tau} = \frac{a}{c} \tag{4.2.4}$$

This however only applies to event A and its vicinity, but how do we know that this applies along the particle's entire world-line?

We consider another frame $\mathcal{S}'$ in which $\mathcal{S}$ has rapidity ρ_S, thus obtained through a boost which we take to be along the particle's acceleration. Since rapidities add, we have that the particle's rapidity in $\mathcal{S}'$ is $\rho' = \rho_A + \rho$ and thus:

$$\frac{d\rho'}{d\tau} = \frac{d\rho_S}{d\tau} + \frac{d\rho}{d\tau} = \frac{d\rho}{d\tau} = \frac{a}{c} \tag{4.2.5}$$

since $\mathcal{S}$ is an inertial frame. So, we see that the time evolution of the rapidity is the same in all inertial frames co-linear with the acceleration. Thus the relation

$$\frac{d\rho}{d\tau} = \frac{a}{c} \tag{4.2.6}$$

applies to the particle's entire motion in any inertial frame.

We can apply this to a rocket undergoing constant linear acceleration. Then we have that:

$$\rho(\tau) = \frac{a\tau}{c} + \text{cnst.} \tag{4.2.7}$$

We can set the constant of integration to zero by considering the particle's rest frame at time $\tau = 0$. Then we find that the particle's speed is:

$$v = c\tanh\left(\frac{a\tau}{c}\right) \tag{4.2.8}$$

Next we wish to relate τ to t in $\mathcal{S}$. We have that:

$$\frac{dt}{d\tau} = \gamma = \cosh\left(\frac{a\tau}{c}\right) \implies t = \frac{c}{a}\sinh\left(\frac{a\tau}{c}\right) \tag{4.2.9}$$

assuming clocks t, τ are synchronized at $t = \tau = 0$. Inserting this into (4.2.8) we reach:

$$\boxed{v(t) = \frac{at}{\sqrt{1 + a^2t^2/c^2}}} \tag{4.2.10}$$

Note that as $t \to \pm\infty$, $v \to \pm c$, an uniformly accelerating particle will seem to approach the speed of light in the infinite time limit. Moreover, we see that:

$$\frac{dv(t)}{dt} = \frac{a}{(1 + a^2t^2/c^2)^{3/2}} \tag{4.2.11}$$

so the acceleration in $\mathcal{S}$ approaches zero as $t \to \infty$, while in the particle's instantaneous rest frame the acceleration remains constant at a.

Finally, we may look at the particle's trajectory. We have that:

$$\frac{dx}{d\tau} = \frac{dx}{dt}\frac{dt}{d\tau} = c\sinh\left(\frac{a\tau}{c}\right) \tag{4.2.12}$$

and thus:

$$x = \frac{c^2}{a}\cosh\left(\frac{a\tau}{c}\right) \tag{4.2.13}$$

where we assume that the particle has position $x = 0$ at $t = 0$. Hence

$$x^2 = \left(\frac{c^2}{a}\right)^2 (1 + \frac{a^2t^2}{c^2}) \iff \boxed{x^2 - c^2t^2 = \frac{c^4}{a^2}} \tag{4.2.14}$$

The particle undergoes hyperbolic motion.

Note that $ds^2 = x^2 - c^2t^2$ is just the space-time interval between the events $(t = 0, x = 0)$ and (t, x). This suggests that a four-vector formulation of this problem. We have that:

$$\mathsf{X} = \frac{c^2}{a}(\cosh\rho, \sinh\rho) \implies \dot{\mathsf{A}} = \frac{a^2}{c^2}\mathsf{U} \tag{4.2.15}$$

Now, for a particle moving with constant acceleration then:

$$0 = \frac{d}{d\tau}(a^2) = \frac{d}{d\tau}(\mathsf{A}\cdot\mathsf{A}) = 2\mathsf{A}\cdot\dot{\mathsf{A}} \propto \mathsf{A}\cdot\mathsf{U} \tag{4.2.16}$$

so the 4-acceleration and 4-velocity are orthogonal.

4.3 Central forces

In the case of a central force, $\mathbf{f} = f(r)\hat{\mathbf{r}}$, we can define the 3-angular momentum as:

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \tag{4.3.1}$$

As in classical mechanics, angular momentum is conserved:

$$\dot{\mathbf{L}} = \mathbf{v} \times \mathbf{p} + \mathbf{r} \times \mathbf{f} = 0 \tag{4.3.2}$$

Consequently, adopting polar coordinates so that $\mathbf{p} = \gamma m(\dot{r}, r\dot{\phi}) \equiv (p_r, \gamma m r\dot{\phi})$, we find:

$$L = \gamma m r^2 \dot{\phi} \iff \frac{L}{mr^2} = \frac{d\phi}{d\tau} \tag{4.3.3}$$

This relates the angular momentum of a particle in some frame to the derivative of the angular position of the particle with respect to proper time.

Now using the energy-momentum relation with $\mathbf{p} = (p_r, \gamma m r\dot{\phi})$, we find that:

$$p_r^2 = \frac{E^2}{c^2} - \frac{L^2}{r^2} - m^2c^2 \tag{4.3.4}$$

Now define the potential energy due to $\mathbf{f}$ as:

$$V = -\int_{\mathcal{O}}^{\mathbf{r}} \mathbf{f} \cdot d\mathbf{r} \tag{4.3.5}$$

Conservation of energy then requires that:

$$E_{tot} \equiv \gamma mc^2 + V = \text{cnst.} \iff p_r^2c^2 + \frac{c^2L^2}{r^2} + m^2c^4 = (\varepsilon - V)^2 \tag{4.3.6}$$

Now:

$$\frac{dr}{d\tau} = \frac{dr}{dt}\frac{dt}{d\tau} = \frac{p_r}{m} \tag{4.3.7}$$

can be substituted into (4.3.6) to get the radial kinetic energy:

$$\frac{1}{2}m\left(\frac{dr}{d\tau}\right)^2 = \frac{(\varepsilon - V)^2 - m^2c^4 - L^2\frac{c^2}{r^2}}{2mc^2} \tag{4.3.8}$$

$$= \varepsilon_{eff} - V_{eff} \tag{4.3.9}$$

where

$$\varepsilon_{eff} = \frac{\varepsilon^2 - m^2c^4}{2mc^2} \tag{4.3.10}$$

$$V_{eff} = \frac{2\varepsilon V - V^2}{2mc^2} + \frac{L^2}{2mr^2} \tag{4.3.11}$$

For a central potential $V(r) = -\frac{\alpha}{r}$:

$$V_{eff} = \frac{-2\alpha\epsilon/r - \alpha^2/r^2}{2mc^2} + \frac{L^2}{2mr^2} \tag{4.3.12}$$

$$= \frac{1}{2mc^2}\left(\frac{L^2c^2 - \alpha^2}{r^2} - \frac{2\alpha\epsilon}{r}\right) \tag{4.3.13}$$

$$= \frac{1}{2mc^2}\left(\frac{(L^2 - L_c^2)c^2}{r^2} - \frac{2\alpha\epsilon}{r}\right) \tag{4.3.14}$$

where we defined $L_c = \frac{\alpha}{c}$. The first term presents dominates at very small r and can be either attractive or repulsive, while the second gives an attractive potential at large r. In the regime where $L > L_c$ and $\varepsilon_{eff} > 0$, then we have stable bound orbits, and we have that:

$$m\frac{d^2r}{d\tau^2}\frac{dr}{d\tau} = -\frac{dV_{eff}}{d\tau} = -\frac{dV_{eff}}{dr}\frac{dr}{d\tau} \tag{4.3.15}$$

$$\iff m\frac{d^2r}{d\tau^2} = -\frac{dV_{eff}}{dr} \tag{4.3.16}$$

$$\iff \frac{d^2r}{d\tau^2} = \tag{4.3.17}$$

4.4 Energy and momentum relations

We begin by justifying our definitions for the energy $E = \gamma(v)m_0\mathbf{v}$ and momentum $p = \gamma(v)m_0\mathbf{v}$.

We consider a general elastic collision between two identical particles (elastic meaning that the rest masses are left unchanged). We choose a frame F such that the two particles have opposite velocities, and orient our axes so that the x-axis bisects the angle of collision, thus ensuring that P^1 is conserved.

We now consider two frames, one moving along the $-x$ direction, following the right particle, and another moving along the $+x$ direction, following the left particle. Let their relative speed be v.

From the first frame's point of view, the right particle doesn't move along the x-axis, only along the y-axis (say with speed u), while the left particle moves along the x-axis with speed v, as well as along the y-axis (say with speed u'). By symmetry, from the second frame's point of view the speeds are exactly the same, but just with reversed roles.

We propose that there is a quantity $\mathbf{p} = \alpha(v)m_0\mathbf{v}$, known as momentum, is conserved in this collision, and investigate whether or not it exists. In the first frame, we see:

$$2\alpha(u)m_0u = 2\alpha(w)m_0u' \implies \frac{\alpha(w)}{\alpha(u)} = \frac{u}{u'} \tag{4.4.1}$$

Lorenz boosting to the second frame, we get $u' = \frac{u}{\gamma(v)}$ and thus:

$$\alpha(w) = \gamma(v)\alpha(u) \tag{4.4.2}$$

Finally, we have that $w^2 = v^2 + (u')^2 = v^2 + u^2 - u^2v^2/c^2$. Setting $\alpha(v) = \gamma(v)$ in general we see that (4.4.2) is satisfied. Therefore, we should have that:

$$\mathbf{p} = \gamma(v)m_0\mathbf{v} \tag{4.4.3}$$

We have yet to consider what happens when the collision involves photons which are massless. We begin by using Planck's relations for photons $E = h\nu$ and $p = h\nu/c$. We consider a mass decaying into two photons. In the mass' rest frame, the photons each have frequency ν, while in some frame moving with speed v to the right, the photons have frequencies ν_1 and ν_2 as shown.

If energy and momentum are to be conserved, in the rest frame:

$$E = 2h\nu, \qquad p = 0 \tag{4.4.4}$$

while in the moving frame:

$$E' = h(\nu_1 + \nu_2), \qquad p' = \frac{h}{c}(\nu_2 - \nu_1) \tag{4.4.5}$$

We now use the longitudinal Doppler equation to relate ν_1 and ν_2:

$$\nu_{2,1} = \sqrt{\frac{1 \pm v/c}{1 \mp v/c}}\nu \tag{4.4.6}$$

$$\implies \nu_1 + \nu_2 = 2\gamma\nu,\ \nu_2 - \nu_1 = 2\gamma\frac{v}{c}\nu \tag{4.4.7}$$

Plugging these into (4.4.5) gives:

$$E' = \gamma E, \qquad p' = \gamma E\frac{v}{c^2} \tag{4.4.8}$$

We now resort to the correspondence principle, our result from Special relativity should reproduce Classical results in the limit $\frac{v}{c} \to 0$. Since in classical mechanics we expect $E' - E = \frac{1}{2}m_0v^2$, we should have:

$$E(\gamma - 1) = \frac{1}{2}m_0v^2 \implies E = mc^2 \tag{4.4.9}$$

finally giving the desired relations:

$$\boxed{E = \gamma mc^2, \quad p = \gamma mv} \tag{4.4.10}$$

4.5 Conservation laws

For a system of N particles with 4-momenta P_i, we define the collective total 4-momentum to be:

$$\boxed{\mathsf{P}(t = t_0) = \sum_i \mathsf{P}_i(t = t_0)} \tag{4.5.1}$$

We have to specify the time at which the sum is taken since in general 4-vectors represent different events in different frames. Here t_0 is the time in the frame in which we are measuring the total 4-momentum. By this definition, in a different frame we must have

$$\mathsf{P}(t' = t'_0) = \sum_i \mathsf{P}_i(t' = t'_0) \tag{4.5.2}$$

However, due to the loss of simultaneity, it is not immediate that one can always find a Lorentz boost Λ such that $\mathsf{P}(t' = t'_0) = \Lambda\mathsf{P}(t = t_0)$. Indeed if the particles have different velocities and don't move as a rigid body then in general $\mathsf{P}_i(t' = t'_0) \neq \Lambda\mathsf{P}_i(t = t_0)$, the individual 4-momenta are not transforms of each other.

If we want the total 4-momentum to be an actual 4-vector that transforms accordingly, then we need a new axiom, the conservation of momentum. Let:

$$\mathsf{P}_{AA} = \text{ 4-momentum in frame A at simultaneous times in frame A} \tag{4.5.3}$$
$$\mathsf{P}_{AB} = \text{ 4-momentum in frame A at simultaneous times in frame B} \tag{4.5.4}$$
$$\mathsf{P}_{BB} = \text{ 4-momentum in frame B at simultaneous times in frame B} \tag{4.5.5}$$

If the conservation of momentum is satisfied then we must have $\mathsf{P}_{AA} = \mathsf{P}_{AB}$, and thus

$$\mathsf{P}_{BB} = \Lambda\mathsf{P}_{AB} = \Lambda\mathsf{P}_{AA} \tag{4.5.6}$$

as desired.

If a sum of 4-vectors evaluated at space-like events (4.5.7a)
is conserved, then this sum is also a 4-vector. (4.5.7b)

It immediately follows that if a 4-vector is conserved in one frame, then it is conserved in all frames.

We prove one final result:

If one component of a 4-vector is conserved in (4.5.8a)
all frames, then the entire 4-vector is conserved. (4.5.8b)

To begin, note that if a component of a 4-vector is null in zero frames, then the entire 4-vector must be zero. Indeed if one of the spatial components is zero in all frames, then by rotations we see that all spatial components must be zero. If the time component is zero in all frames, but at least one spatial component is not, then we can Lorentz boost along that component to make the time component non-zero, a contradiction. Hence all components of the four-vector must be zero.

Suppose P has a component P^μ that is conserved so that $P^\mu = P'^\mu$. Then letting $\mathsf{Q} = \mathsf{P}' - \mathsf{P}$, and applying the lemma we have proven, we see that $\mathsf{Q} = 0$, and thus the entire 4-vector P is conserved.

4.6 Relativistic collisions

We can now use the tools we have developed on conservation laws to examine a plethora of relativistic collisions.

Radioactive decay/absorption

Suppose a particle of mass M decays into two smaller particles of masses m_1 and m_2. In the rest frame of the initial particle, the four-momentum of M reads $\mathsf{P}_1 = (Mc, 0, 0, 0)$, while for the final two particles it is $\mathsf{P}_2 = (E_1/c, p_1, 0, 0)$ and $\mathsf{P}_3 = (E_2/c, p_2, 0, 0)$. Conservation of 4-momentum implies that:

$$E_1 + E_2 = Mc^2, \qquad p_1 = -p_2 \tag{4.6.1}$$

The energy-momentum equivalence relation also implies that:

$$E_1^2 - p_1^2c^2 = m_1^2c^4, \qquad E_2^2 - p_1^2c^2 = m_2^2c^4 \tag{4.6.2}$$

$$\iff (E_1 - E_2)(E_1 + E_2) = (m_1^2 - m_2^2)c^4 \tag{4.6.3}$$

$$\iff E_1 - E_2 = \frac{m_1^2 - m_2^2}{M}c^2 \tag{4.6.4}$$

$$\iff \boxed{E_1 = \frac{m_1^2 - m_2^2 + M^2}{2M}c^2} \tag{4.6.5}$$

Suppose one of the particles is a photon so that $m_1 = 0$. Let $E_0 = Mc^2 - m_2c^2$ be the change in rest mass energy. Then:

$$E_1 = \frac{M^2 - m_2^2}{2M}c^2 = \left(1 - \frac{E_0}{2Mc^2}\right)E_0 \tag{4.6.6}$$

so the energy of the photon is slightly smaller than the rest energy change, with:

$$E_1 - E_0 = -\frac{E_0^2}{2Mc^2} \tag{4.6.7}$$

known as the recoil energy reducing the photon energy. The recoil energy is required to recoil the mass m_2 as required by conservation of momentum.

If instead we have a mass m_2 strike a mass m_1 thus forming a larger mass M, then one can easily find through the same process as the case of emission that:

$$\boxed{E_1 = \frac{-m_1^2 - m_2^2 + M^2}{2M}c^2} \tag{4.6.8}$$

Two-particle decay

Suppose a particle of mass M decays into several smaller particles. We have that:

$$\mathsf{P} = \sum_i \mathsf{P}_i \tag{4.6.9}$$

and thus

$$M^2c^4 = \left(\sum_i E_i\right)^2 - \left(\sum_i \mathbf{p}_i\right)\cdot\left(\sum_i \mathbf{p}_i\right)c^2 \tag{4.6.10}$$

If we only have two decay products then:

$$\mathsf{P} = \mathsf{P}_1 + \mathsf{P}_2 \implies M^2c^2 = m_1^2c^2 + m_2^2c^2 + 2\mathsf{P}_1\cdot\mathsf{P}_2 \tag{4.6.11}$$

Clearly $\mathsf{P}_1\cdot\mathsf{P}_2 = \gamma(u)m_1m_2c^2$ (evaluate this product in the rest frame of one of the particles) where u is the relative speed of one decay product relative to the other. Hence:

$$M^2 = m_1^2 + m_2^2 + 2\gamma(u)m_1m_2 \tag{4.6.12}$$

If one is able to measure the outgoing particles' masses and relative speeds, then we can trace back to the original mass.

Threshold energy and the CM frame

Suppose we take a particle of mass m with energy E, momentum $\mathbf{p}$ and collide it with another particle of mass M with the goal of creating new particles.

We can consider this from the center of mass frame where $\mathsf{P}_{CM} = (E_{CM}/c, \mathbf{0})$, while in the laboratory frame $\mathsf{P} = (E/c + Mc, \mathbf{p})$. Thus:

$$E_{CM}^2 = (E + Mc^2)^2 - p^2c^2 = m^2c^4 + M^2c^4 + 2EMc^2 \tag{4.6.13}$$

Our goal is to find the minimum E, known as **threshold energy**, such that the collision may create several particles of total rest mass $\sum_i m_i$. Clearly, this is achieved when all the particles move with momentum p in the lab frame, and thus no momentum in the CM frame. In this case $E_{CM} = \sum_i m_ic^2$ which when substituted into (4.6.13) gives the threshold energy:

$$\boxed{E_{th} = \frac{\left(\sum_i m_i\right)^2 - m^2 - M^2}{2M}c^2} \tag{4.6.14}$$

It is also useful to know what is the relative velocity between the CM frame and lab frame. Suppose we have a system with momentum $\mathbf{p}$ and energy E in the lab frame. WLOG we can align our x-axis with $\mathbf{p}$, and thus Lorentz boost to the CM frame:

$$E_{CM} = \gamma(v)(E - pv), \qquad 0 = \gamma(v)(vE/c^2 - p) \tag{4.6.15}$$

the latter of which gives $v = \frac{pc^2}{E}$ and hence $E_{CM} = \gamma\frac{E^2-p^2c^2}{E}$.

Three-body decay

We now consider a particle of mass M decaying into three products of masses m_1, m_2, m_3. We have that:

$$\mathsf{P} = \mathsf{P}_1 + \mathsf{P}_2 + \mathsf{P}_3 \tag{4.6.16}$$

Now a useful trick when solving collisions problems is squaring both sides of the momentum conservation law.

$$(\mathsf{P}-\mathsf{P}_3)^2=(\mathsf{P}_1+\mathsf{P}_2)^2 \implies M^2c^2+m_3c^2-2\mathsf{P}\cdot\mathsf{P}_3=m_1^2c^2+m_2c^2+2\mathsf{P}_1\cdot\mathsf{P}_2 \quad (4.6.17)$$

Note that the result is symmetric in m, M reflecting the fact that while in our derivation m was made to collide with M, the opposite picture may also be taken.

Elastic collisions

In an elastic collision the colliding particles do not undergo any change in mass. This alone allows us to derive an interesting result with a classical analogue. Suppose two particles with 4-momenta P and Q collide elastically, outgoing with 4-momenta P' and Q'. Conservation of momentum implies that

$$\mathsf{P}+\mathsf{Q}=\mathsf{P}'+\mathsf{Q}' \quad (4.6.18)$$

$$\iff \mathsf{P}^2+\mathsf{Q}^2+2\mathsf{P}\cdot\mathsf{Q}=\mathsf{P}'^2+\S Q'^2+2\mathsf{P}'\cdot\mathsf{Q}' \quad (4.6.19)$$

$$\iff \boxed{\mathsf{P}\cdot\mathsf{Q}=\mathsf{P}'\cdot\mathsf{Q}'} \quad (4.6.20)$$

Consequently, since $\mathsf{P}\cdot\mathsf{Q}\propto\gamma_u$ where u is the relative velocities of the particles, we see that the particles will have the same relative velocity before and after the collision. Note that the same result holds in classical mechanics.

Consider two identical particles of mass m colliding. We adopt the rest frame of one of the particles and orient our axes so that the x-axis points along the collision line.

We find that before the collision the particles have 4-momenta:

$$\mathsf{P}_1=(\gamma_u mc,\gamma_u mu,0,0) \quad (4.6.21)$$

$$\mathsf{P}_2=(mc,0,0,0) \quad (4.6.22)$$

while after the collision they are:

$$\mathsf{P}_3=(\gamma_v mc,\gamma_v mv\cos\theta_1,\gamma_v mv\sin\theta_1,0) \quad (4.6.23)$$

$$\mathsf{P}_4=(\gamma_w mc,\gamma_w mw\cos\theta_1,-\gamma_w mw\sin\theta_1,0) \quad (4.6.24)$$

Conservation of momentum then yields:

$$\gamma_u+1=\gamma_v+\gamma_w \quad (4.6.25)$$

$$\gamma_u\mathbf{u}=\gamma_v\mathbf{v}+\gamma_w\mathbf{w} \quad (4.6.26)$$

The second gives:

$$\gamma_u^2u^2=\gamma_v^2v^2+\gamma_w^w w^2+2\gamma_v\gamma_w\mathbf{v}\cdot\mathbf{w} \quad (4.6.27)$$

and substituting the first into the above we find

$$(\gamma_v+\gamma_w-1)^2u^2=\gamma_v^2v^2+\gamma_w^w w^2+2\gamma_v\gamma_w\mathbf{v}\cdot\mathbf{w} \quad (4.6.28)$$

and using the relation $\gamma_v^2 v^2 = (\gamma_v^2 - 1)c^2$ we find:

$$(\gamma_v + \gamma_w - 1)^2 c^2 - c^2 - \gamma_v^2 v^2 - \gamma_w^w w^2 = 2\gamma_v \gamma_w v w \cos\theta \tag{4.6.29}$$

$$\implies 2c^2(\gamma_v - 1)(\gamma_w - 1) = 2\gamma_v \gamma_w v w \cos\theta \tag{4.6.30}$$

$$\implies \boxed{\cos\theta = \frac{(\gamma_v - 1)(\gamma_w - 1)}{\gamma_v \gamma_w v w} c^2 = \sqrt{\frac{\gamma_v - 1}{\gamma_v + 1} \frac{\gamma_w - 1}{\gamma_w + 1}}} \tag{4.6.31}$$

This gives the angle between the outgoing elastically collided particles. In the low speed limit the particles leave at right angles to each other, and as we increase the speeds θ decreases.

Compton scattering

5

Covariant electromagnetism

5.1 Remarks on relativistic waves

5.2 The Continuity equation and 4-current

Electric charge is locally conserved, this is expressed using the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \tag{5.2.1}$$

If it were possible to establish $\mathsf{J} = (\rho c, \mathbf{J})$ as a 4-vector, then one could neatly write the continuity equation in a Lorentz covariant form: [1] $\Box \cdot \mathsf{J} \equiv \partial_\mu J^\mu = 0$

Consider two frames $\mathcal{S}$ and $\mathcal{S}'$ moving with relative velocity $\mathbf{u}$. In frame $\mathcal{S}$ a finite region of charge density ρ moves with velocity $\mathbf{v}$ to the right as shown:

Due to the Lorentz invariance of charge, we must have that the same amount of charge must be contained within an infinitesimal volume, so that:

$$\rho d\mathbf{r} = \rho' d\mathbf{r}' \tag{5.2.2}$$

Now letting w be the speed of the charge volume in $\mathcal{S}'$ then clearly $\gamma_w = \gamma_v \gamma_u (1 + \frac{\mathbf{u} \cdot \mathbf{v}}{c^2})$ by velocity-addition. Hence:

$$d\mathbf{r} = \frac{d\mathbf{r}_0}{\gamma_v} \implies d\mathbf{r}' = \frac{\gamma_v}{\gamma_w} d\mathbf{r} = \frac{d\mathbf{r}}{\gamma_u (1 + \mathbf{u} \cdot \mathbf{v}/c^2)} \tag{5.2.3}$$

which gives:

$$\rho' = \gamma_u \left(\rho + \frac{\mathbf{J} \cdot \mathbf{u}}{c^2} \right) \tag{5.2.4}$$

as desired. We now make use of the definition $\mathbf{J} = \rho \mathbf{v}$ and $\mathbf{J}'_{\|} = \rho' \mathbf{w}_{\|}$ to get the transformation of parallel components:

$$\mathbf{J}'_{\|} = \gamma_u \left(\rho + \frac{\mathbf{J} \cdot \mathbf{u}}{c^2} \right) \mathbf{w}_{\|} = \gamma_u \left(\rho + \rho \frac{\mathbf{v} \cdot \mathbf{u}}{c^2} \right) \frac{\mathbf{u} + \mathbf{v}}{1 + \mathbf{u} \cdot \mathbf{v}/c^2} = \gamma_u \rho (\mathbf{u} + \mathbf{v}) \tag{5.2.5}$$

[1] Lorentz covariant means that it makes no reference to frame coordinates, sort of like how Newton's laws in vector form are Galilean covariant as they don't make reference to spatial coordinates

which gives:

$$\mathbf{J}'_{\parallel} = \gamma_u(\mathbf{J} + \rho\mathbf{u}) \tag{5.2.6}$$

Finally,

$$\mathbf{J}'_{\perp} = \gamma_u\left(\rho + \frac{\mathbf{J}\cdot\mathbf{u}}{c^2}\right)\mathbf{w}_{\perp} = \gamma_u\left(\rho + \rho\frac{\mathbf{v}\cdot\mathbf{u}}{c^2}\right)\frac{\mathbf{v}}{\gamma_u(1+\mathbf{u}\cdot\mathbf{v}/c^2)} = \rho\mathbf{v} \tag{5.2.7}$$

which gives:

$$\mathbf{J}'_{\perp} = \mathbf{J}_{\perp} \tag{5.2.8}$$

It follows that $(\rho c, \mathbf{J})$ transforms as a 4-vector which we call the 4-current. We could have also noted that $\mathsf{J} = \rho_0 \mathsf{U}$ where ρ_0 is the rest charge density, a Lorentz scalar. The continuity equation takes the form:

$$\Box \cdot \mathsf{J} = 0 \tag{5.2.9}$$

5.3 E and B, two sides of the same coin

Our discussion on charges and currents suggest that there is an interplay between charge distributions and current distributions, which themselves produce electric and magnetic fields. As Lorentz transforming charges produce currents and vice versa, one should expect that Lorentz transforming electric fields should produce magnetic fields too.

Consider in some frame $\mathcal{S}$ a neutral wire carrying a current I (made of moving positive charges). If we place a test charge at some radial distance r with initial speed v along the wire, then one would expect the force on it to be a purely magnetic Lorentz force:

$$F_{mag} = -\frac{qv\mu_0 I}{2\pi r} \tag{5.3.1}$$

Let's now boost to the test charge's rest frame $\mathcal{S}'$. Now the positive charge density will be $\rho_+ = \rho$ in $\mathcal{S}$ and hence $\rho'_+ = \gamma_v\rho\left(1 - \frac{uv}{c^2}\right)$ in $\mathcal{S}'$ while the negative charge density will be $\rho_- = -\rho$ in $\mathcal{S}$ and hence $\rho'_- = -\gamma_v\rho$ in $\mathcal{S}'$. The test particle will thus experience no magnetic force but an electrostatic force due to a net charge density $\rho' = \gamma_v\rho\frac{uv}{c^2}$. If the wire has cross-section A then the electric field produced will be:

$$F'_{el} = -\gamma_v\frac{q\rho uvA}{2\pi c^2\varepsilon_0 r} = -\gamma_v\frac{q\mu_0\rho uvA}{2\pi r} \tag{5.3.2}$$

We can transform this form in the original frame to find:

$$F_{el} = -\frac{q\mu_0\rho uvA}{2\pi r} \tag{5.3.3}$$

Recall that if the wire has current I and cross-section A then $I = nAeu = \rho Au$ where n is the charge carrier density and e the electron charge. Therefore the above result may be rewritten as:

$$F_{el} = -\frac{qv\mu_0 I}{2\pi r} = F_{mag} \tag{5.3.4}$$

which is precisely the magnetic force we calculated earlier! In hindsight there was no real need to define a magnetic force, all of this could be calculated using Lorentz contraction and Coulomb's law.

5.4 Gauge invariance

What is a gauge?

We now seek to find a more general law of transformation between the electric and magnetic fields. To do so we must look at the gauge invariance of Maxwell's equations.

$$\nabla \cdot \mathbf{E} = \rho \tag{5.4.1}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{5.4.2}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{5.4.3}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} \tag{5.4.4}$$

From the second equation and the Hemholtz decomposition theorem we see that we may write $\mathbf{B} = \nabla \times \mathbf{A}$ where $\mathbf{A}$ is a vector potential. It then follows that:

$$\nabla \times \mathbf{E} = -\nabla \times \frac{\partial \mathbf{A}}{\partial t} \implies \nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0 \tag{5.4.5}$$

which means the electric and magnetic field may be written as functions of the scalar and vector potentials:

$$\boxed{\mathbf{B} = \nabla \times \mathbf{A}, \qquad \mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}} \tag{5.4.6}$$

These equations have a hidden symmetry, known as a Gauge invariance, which follows from the fact that the curl of a gradient is null. Consequently, suppose we perform the transformation $\mathbf{A}' \mapsto \mathbf{A} + \nabla\chi$ for some well-behaved χ:

$$\nabla \times \mathbf{A}' = \nabla \times \mathbf{A} + \nabla \times (\nabla\chi) = \mathbf{E} \tag{5.4.7}$$

We therefore have an infinite family of possible $\mathbf{A}$ for a given $\mathbf{A}$. This is somehow reminiscent of how an indefinite integral has infinitely many possible values due to the fact that the derivative of a constant is zero. We can extend this argument to $\mathbf{E}$:

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} - \frac{\partial(\nabla\chi)}{\partial t} \tag{5.4.8}$$

so if we want this gauge invariance to apply to $\mathbf{E}$ then we need $\phi \mapsto \phi - \frac{\partial\chi}{\partial t}$. With this choice then:

$$\mathbf{E} = -\nabla\phi + \nabla\frac{\partial\chi}{\partial t} - \frac{\partial \mathbf{A}}{\partial t} - \frac{\partial(\nabla\chi)}{\partial t} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \tag{5.4.9}$$

as desired.

To summarize, our definitions of **E** and **B** are invariant under gauge transformations:

$$\boxed{\phi \mapsto \phi + \frac{\partial \chi}{\partial t}, \qquad \mathbf{A} \mapsto \mathbf{A} - \nabla \chi} \tag{5.4.10}$$

These transformations can be written more succintly as:

$$(\phi/c, \mathbf{A}) \mapsto (\phi/c - \frac{1}{c}\frac{\partial \chi}{\partial t}, \mathbf{A} + \nabla \chi) \tag{5.4.11}$$

which suggests postulating that $A^\mu = (\phi/c, \mathbf{A})$ is a 4-vector. If this is the case then a gauge transformation can be written as:

$$A^\mu \mapsto A^\mu + \partial^\mu \chi \tag{5.4.12}$$

One very famous gauge that is often used in classical electromagnetism is the Coulomb gauge:

$$\nabla \cdot \mathbf{A} = 0 \tag{5.4.13}$$

With this gauge one obtains the homogeneous wave equations:

$$\nabla^2 \mathbf{A} - \frac{1}{c^2}\frac{\partial \mathbf{A}}{\partial t^2} = 0 \tag{5.4.14}$$

as can be easily verified. Unfortunately this gauge is incompatible with special relativity because it does not treat time and space on equal footing (it is not Lorentz covariant). It would be nice to have a gauge condition that is manifestly covariant.

The Lorentz gauge

With this in mind, we try to formulate Ampere-Maxwell's law using the vector potential:

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \frac{1}{c^2}\frac{\partial(\nabla \phi)}{\partial t} \tag{5.4.15}$$

$$\iff \nabla\left(\nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t}\right) + \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J} \tag{5.4.16}$$

Note that $\nabla \cdot \mathbf{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t} \equiv \partial_\mu A^\mu$. It would be nice to set this equal to zero, so we define a new gauge known as the Lorentz gauge:

$$\boxed{\Box \cdot \mathsf{A} = 0} \tag{5.4.17}$$

Note that this finally shows that A^μ is a 4-vector, since its dot product with the 4-gradient gives a Lorentz scalar.

Also, it is always possible to find a Lorentz gauge for a given **E**, **B**. Indeed, suppose we have some 4-potential A^μ such that $\partial_\mu A^\mu = f$. Then if we perform some gauge transformation $A'^\mu = A^\mu + \partial^\mu \chi$ we find:

$$\partial_m u A'^\mu = \partial_\mu A^\mu + \Box^2 \chi \tag{5.4.18}$$

For this to be zero we require $\Box^2 \chi = -f$. Due to the existence and uniqueness theorem this can always be done so one can always use the Lorentz gauge.

With this in mind we get that:

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \mu_0 \mathbf{J} \implies \Box^2 \mathbf{A} = \mu_0 \mathbf{J} \tag{5.4.19}$$

Knowing that $(\phi/c, \mathbf{A})$ and $(\rho c, \mathbf{J})$ are 4-vector we should expect a very similar equation to hold for ρ. We can use Gauss's law to write:

$$\nabla^2 \phi + \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = \nabla^2 \phi - \frac{1}{c^2}\frac{\partial^2 \phi}{\partial t^2} = \rho/\varepsilon_0 \tag{5.4.20}$$

$$\iff \nabla^2(\phi/c) - \frac{1}{c^2}\frac{\partial^2 \phi/c}{\partial t^2} = \mu_0 \rho c \implies \Box^2 \phi = \mu_0 \rho \tag{5.4.21}$$

We can combine $\Box^2 \mathbf{A} = \mu_0 \mathbf{J}$ and $\Box^2 \mathbf{A} = \mu_0 \rho$ into a single, manifestly covariant equation:

$$\Box^2 \mathsf{A} = \mu_0 \mathsf{J} \tag{5.4.22}$$

We have not yet proven that it is possible to find a Lorentz gauge for all possible electromagnetic configurations. We need to find a gauge transformation that reduces any given 4-potential to a Lorentz gauge.

Suppose that we are given some potential A_μ which does not satisfy the Lorentz gauge condition: $\partial_\mu A^\mu = \varphi \neq 0$ where φ is some function. When we perform a gauge transformation, we find that that the new gauge must satisfy $\partial_\mu A^\mu + \Box^2 \chi = \varphi$. For the Lorentz condition to hold we require $\Box^2 \chi = \varphi$:

$$\frac{1}{c^2}\frac{\partial^2 \chi}{\partial t^2} - \nabla^2 \chi = \varphi \tag{5.4.23}$$

But the wave-equation has an existence and uniqueness theorem, thus given the necessary boundary conditions this wave-equation always has a solution.

5.5 Making Electromagnetism covariant

The electromagnetic field tensor

With our development of the 4-potential we now seek to write Maxwell's equations in manifestly covariant form. To do so we will need a quantity which encodes both $\mathbf{E}$ and $\mathbf{B}$ and that follows Lorentzian transformation laws.

Clearly this cannot be a 4-vector since we have a total of 6 electromagnetic field components. The next logical step is a 4-tensor $F^{\mu\nu}$ which transforms as:

$$F'^{\mu\nu} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta F^{\alpha\beta} \iff \mathbb{F}' = \Lambda \mathbb{F} \Lambda^T \tag{5.5.1}$$

This is easily done by We can define the following rank-2 tensor:

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{5.5.2}$$

known as the electromagnetic field tensor. One very important property of this tensor is that it is anti-symmetric. Consequently $F^{\mu\mu} = 0$.

Note also that $A^\mu \to A^\mu + \Box\chi$ then:

$$F^{\mu\nu} \to \partial^\mu(A^\nu + \Box\chi) - \partial^\nu(A^\mu + \Box\chi) = F^{\mu\nu} \tag{5.5.3}$$

so the electromagnetic field tensor is gauge invariant as one would require for it to encode information about **E** and **B**.

Now we know that $F^{\mu\nu}$ will definitely include the electric and magnetic fields as we are taking derivatives of the potentials. Indeed:

$$F^{i0} = \frac{1}{c}\partial^i\phi - \frac{1}{c}\partial^0\mathbf{A} = E^i \implies F^{0i} = -E_i/c \tag{5.5.4}$$

Similarly:

$$F^{12} = -\frac{\partial A_y}{\partial x} + \frac{\partial A_x}{\partial y} = -B_3 \tag{5.5.5}$$

We can cycle through the indices and find that $F^{13} = B_2$ and $F^{23} = -B_1$. In general it is easy to see that:

$$B_i = \frac{1}{2}\epsilon_{ijk}F^{jk}, \quad E^i = cF^{i0} \tag{5.5.6}$$

Thus:

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \tag{5.5.7}$$

The Electromagnetic field equations

Immediately we see that:

$$\partial_\mu F^{\mu\nu} = \partial_\mu\partial^\mu A^\nu - \partial_\mu\partial^\nu A^\mu = \Box^2 A^\nu \tag{5.5.8}$$

so using (5.4.22) we find that:

$$\partial_\mu F^{\mu\nu} = \mu_0 J^\nu \tag{5.5.9}$$

Also, we see that due to the antisymmetry of the electromagnetic field tensor the following must also hold:

$$\partial_{[\alpha}F_{\beta\gamma]} \equiv \partial_\alpha F_{\beta\gamma} + \partial_\gamma F_{\alpha\beta} + \partial_\beta F_{\gamma\alpha} = 0 \tag{5.5.10}$$

known as the Bianchi identity. It is easy to see that this reproduces the homogeneous Maxwell equations.

We can write (5.5.10) in another way by introducing the dual electromagnetic field tensor:

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} \tag{5.5.11}$$

It is then easy to see that due to the anti-symmetry of the Levi-Civita 4-tensor:

$$\partial_\mu \tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}\partial_\mu F_{\alpha\beta} \tag{5.5.12}$$

$$= \frac{1}{6}\epsilon^{\mu\nu\alpha\beta}(\partial_\mu F_{\alpha\beta} + \partial_\mu F_{\alpha\beta} + \partial_\mu F_{\alpha\beta}) \tag{5.5.13}$$

$$= \frac{1}{6}\epsilon^{\mu\nu\alpha\beta}(\partial_\mu F_{\alpha\beta} + \partial_\beta F_{\mu\alpha} + \partial_\alpha F_{\beta\mu}) \tag{5.5.14}$$

We recognize that the factor in parenthesis must vanish, so we find:

$$\boxed{\partial_\mu \tilde{F}^{\mu\nu} = 0} \tag{5.5.15}$$

Maxwell's equations have thus been reduced to two manifestly covariant equations:

$$\boxed{\partial_\mu F^{\mu\nu} = \mu_0 J^\mu, \qquad \partial_\mu \tilde{F}^{\mu\nu} = 0} \tag{5.5.16}$$

5.6 Lorentz transforming the Lorentz force

Manifestly covariant Lorentz force

In classical electromagnetism we define the electric and magnetic fields as vector fields embedded in space which act on a charge q with a Lorentz force:

$$\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{5.6.1}$$

We can write this as:

$$f^i = q(E^i + \epsilon^{ijk}v_j B_k) \tag{5.6.2}$$

$$= q(cF^{i0} + \epsilon^{ijk}v_k B_i) \tag{5.6.3}$$

$$= q(cF^{i0} + F^{ij}v_j) = qF^{i\mu}U_\mu \tag{5.6.4}$$

which suggests writing down more generally that:

$$\boxed{\mathsf{F} = q\mathbb{F}\cdot\mathsf{U} \iff f^\mu = qF^{\mu\nu}U_\nu} \tag{5.6.5}$$

which gives an additional equation:

$$\frac{dE_{en}}{dt} = q\mathbf{v}\cdot\mathbf{E} \tag{5.6.6}$$

where E_{en} is the energy, and not the electric field amplitude. We can make sense of this equation if the Lorentz force is a pure force (which it should be, electromagnetic fields can only accelerate particles), then we see that:

$$\frac{dE}{dt} = \mathbf{v}\cdot\mathbf{f} = \mathbf{v}\cdot\mathbf{E} \tag{5.6.7}$$

E and B transformations

We now use the fact that the electromagnetic field tensor is a tensor to derive the transformation laws of the electric and magnetic fields. We see that:

$$\begin{aligned}E'_x &= F'^{10} = \Lambda^1_\mu \Lambda^0_\nu F^{\mu\nu} \\ &= \Lambda^1_0 \Lambda^0_1 F^{01} + \Lambda^1_1 \Lambda^0_0 F^{10} \\ &= -\beta^2\gamma^2 E_x + \gamma^2 E_x \\ &= E_x\end{aligned}$$

$$\begin{aligned}E'_y &= F'^{20} = \Lambda^2_\mu \Lambda^0_\nu F^{\mu\nu} \\ &= \Lambda^2_2 \Lambda^0_1 F^{21} + \Lambda^2_2 \Lambda^0_0 F^{20} \\ &= -\gamma\beta c B_z + \gamma E_y \\ &= \gamma(E_y - vB_z)\end{aligned}$$

$$\begin{aligned}E'_z &= F'^{30} = \Lambda^3_\mu \Lambda^0_\nu F^{\mu\nu} \\ &= \Lambda^3_3 \Lambda^0_1 F^{31} + \Lambda^3_3 \Lambda^0_0 F^{30} \\ &= \gamma\beta c B_y + \gamma E_z \\ &= \gamma(E_z + vB_y)\end{aligned}$$

$$\begin{aligned}B'_x &= F'^{32} = \Lambda^3_\mu \Lambda^2_\nu F^{\mu\nu} \\ &= \Lambda^3_3 \Lambda^2_2 F^{32} \\ &= B_x\end{aligned}$$

$$\begin{aligned}B'_y &= F'^{13} = \Lambda^1_\mu \Lambda^3_\nu F^{\mu\nu} \\ &= \Lambda^1_1 \Lambda^3_3 F^{13} + \Lambda^1_0 \Lambda^3_3 F^{03} \\ &= \gamma B_y + \beta\gamma E_z/c \\ &= \gamma(B_y + v/c^2 E_z)\end{aligned}$$

$$\begin{aligned}B'_z &= F'^{21} = \Lambda^2_\mu \Lambda^1_\nu F^{\mu\nu} \\ &= \Lambda^2_2 \Lambda^1_0 F^{20} + \Lambda^2_2 \Lambda^1_1 F^{21} \\ &= -\gamma\beta E_y/c + \gamma B_z \\ &= \gamma(B_z - v/c^2 E_y)\end{aligned}$$

Consequently for boosts along the x-axis:

$$\boxed{\begin{aligned}E'_x &= E_x & B'_x &= B_x \\ E'_y &= \gamma(E_y - vB_z) & B'_y &= \gamma(B_y + v/c^2 E_z) \\ E'_z &= \gamma(E_z + vB_y) & B'_z &= \gamma(B_z - v/c^2 E_y)\end{aligned}}$$

These can be generalized to:

$$\boxed{\begin{aligned}\mathbf{E}'_\parallel &= \mathbf{E}_\parallel & \mathbf{B}'_\parallel &= \mathbf{B}_\parallel \\ \mathbf{E}'_\perp &= \gamma(\mathbf{E}_\perp + \mathbf{v}\times\mathbf{B}) & \mathbf{B}'_\perp &= \gamma(\mathbf{B}_\perp - \mathbf{v}\times\mathbf{E}/c^2)\end{aligned}}$$

As we can see, the electric field in one frame morphs into part of the magnetic field in another frame, thus explaining the phenomenon in 5.3, as well as most of the interactions in the natural world.

Electromagnetic radiation

6

In classical electromagnetism it is known that Maxwell's equations allow for electromagnetic waves. We are interested in seeing how such waves can be generated in the first place, how does one produce a changing electric and magnetic field? The answer is accelerating charges.

6.1 The Hemholtz equation

In the Lorentz gauge $\partial_\mu A^\mu = 0$ the inhomogeneous maxwell equations read:

$$\Box^2 A^\mu = \mu_0 J^\mu \iff \left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)A^\mu = \mu_0 J^\mu \tag{6.1.1}$$

Let us take a temporal Fourier transform:

$$A_\mu(\mathbf{x}, t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\tilde{A}_\mu(\mathbf{x}, \omega)e^{-i\omega t},\ J_\mu(\mathbf{x}, t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\tilde{J}_\mu(\mathbf{x}, \omega)e^{-i\omega t} \tag{6.1.2}$$

and substitute into (6.1.1):

$$\int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\tilde{A}_\mu(\mathbf{x}, \omega)e^{-i\omega t} = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi}\mu_0\tilde{J}_\mu(\mathbf{x}, \omega)e^{-i\omega t} \tag{6.1.3}$$

$$\implies \left(\nabla^2 + \frac{\omega^2}{c^2}\right)\tilde{A}_\mu = -\mu_0 J_\mu \tag{6.1.4}$$

The last equation is known as the **Hemholtz equation**, and can be solved using Green's functions. We find that:

$$\left(\nabla^2 + \frac{\omega^2}{c^2}\right)G(\mathbf{x}, \mathbf{x}') = \delta^3(\mathbf{x} - \mathbf{x}') \tag{6.1.5}$$

The Hemholtz equation is spherically symmetric so the solution can only depend on the radial coordinate $r = |\mathbf{x} - \mathbf{x}'|$. Then we claim that the following are Green's functions:

$$G_\pm(r) = -\frac{1}{4\pi}\frac{e^{\pm ikr}}{r},\ r \neq 0 \tag{6.1.6}$$

where $k = \frac{\omega}{c}$. Indeed, we have that:

$$\nabla^2 G_\pm(r) = -\frac{1}{4\pi r}\nabla^2 e^{\pm ikr} + e^{\pm ikr}\nabla^2\Big(-\frac{1}{4\pi r}\Big) + 2(\nabla e^{\pm ikr})\cdot\nabla\Big(-\frac{1}{4\pi r}\Big) \tag{6.1.7}$$

Term by term, we have that:

$$\nabla^2 e^{\pm ikr} = \nabla\cdot(\pm ike^{\pm ikr}\hat{\mathbf{r}}) = \Big(-k^2 \pm \frac{2ik}{r}\Big)e^{\pm ikr} \tag{6.1.8}$$

$$\nabla^2\Big(-\frac{1}{4\pi r}\Big) = \delta^3(\mathbf{r}) \tag{6.1.9}$$

$$(\nabla e^{\pm ikr})\cdot\nabla\Big(-\frac{1}{4\pi r}\Big) = (\pm ike^{\pm ikr}\hat{\mathbf{r}})\cdot\Big(\frac{1}{4\pi r^2}\hat{\mathbf{r}}\Big) \tag{6.1.10}$$

finally giving:

$$\nabla^2 G_\pm(r) = -k^2 G_\pm(r) + \delta^3(\mathbf{r}) \tag{6.1.11}$$

as desired. Consequently the general solution to (6.1.5) is:

$$A_\mu(\mathbf{x},t) = \frac{\mu}{4\pi}\int\frac{d\omega}{2\pi}\int d^3\mathbf{x}'\frac{e^{-i\omega(t-|\mathbf{x}-\mathbf{x}'|/c)}}{|\mathbf{x}-\mathbf{x}'|}\tilde{J}_\mu(\mathbf{x}',\omega) \tag{6.1.12}$$

For reasons that we shall clarify soon we only kept the G_+ green's function. We can define the retarded time as:

$$t_{ret} = t - \frac{|\mathbf{x}-\mathbf{x}'|}{c} \tag{6.1.13}$$

which finally gives the **retarded potential**:

$$A_\mu(\mathbf{x},t) = \frac{\mu}{4\pi}\int d^3\mathbf{x}'\frac{J_\mu(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|} \tag{6.1.14}$$

Surprisingly, our general solution for the 4-potential is quite similar to the stationary 4-current solution (Coulomb and Biot-Savart laws). The only difference is that we must integrate over the 4-current at a retarded time t_{ret} rather than t. This is a consequence of causality: the fact that if a we perturb the 4-current at $(\mathbf{x}', t')$ then an observer at position $\mathbf{x}$ will have to wait $t - t_{ret}$ time to obtain this information. So to the observer the 4-current is as it actually is at (proper) time $t - t_{ret}$.

We now see why the Green's function G_+ could not have been chosen. It would have violated causality, implying that to know the 4-potential at time t one should have knowledge of the 4-current at a later time $t_{adv} = t + \frac{|\mathbf{x}-\mathbf{x}'|}{c}$.

6.2 Retarded and advanced Green's functions

There is another method to derive the advanced and retarded potentials which is quite useful, especially in later courses (e.g. QFT). Instead of finding the Green's functions from the Hemholtz function, we start directly with the wave equation:

$$\left(\nabla^2 - \frac{1}{c}^2 \frac{\partial^2}{\partial t^2}\right) G(\mathbf{r}, t) = \delta^3(\mathbf{r})\delta(t) \tag{6.2.1}$$

where we set $\mathbf{x}' = 0$ and $t' = 0$. We can once again take a Fourier transform, this time both in space and time:

$$G(\mathbf{r}, t) = \int \frac{d\omega d^3\mathbf{k}}{(2\pi)^4} \tilde{G}(\mathbf{k}, t) e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{6.2.2}$$

One may be initially perplexed by the negative sign in the exponent. Relativistically, note that $\mathsf{K} \cdot \mathsf{X} = \omega t - \mathbf{k} \cdot \mathbf{r}$ thus giving the negative sign in the Fourier transform [1]. Physically, this means that we want to decompose our solutions into waves propagating forwards in time, rather than backwards.

The wave-equation now reads:

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) G(\mathbf{r}, t) = \delta^3(\mathbf{r})\delta(t) \tag{6.2.3}$$

$$\implies \int \frac{d\omega d^3\mathbf{k}}{(2\pi)^4} \tilde{G}(\mathbf{k}, \omega) \Big(-k^2 + \frac{\omega^2}{c^2} \Big) e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} = \int \frac{d\omega d^3\mathbf{k}}{(2\pi)^4} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{6.2.4}$$

$$\implies \qquad \tilde{G}(\mathbf{k}, \omega) = -\frac{1}{k^2 - \omega^2/c^2} \tag{6.2.5}$$

and reverting the Fourier transform:

$$G(\mathbf{r}, t) = -\int \frac{d\omega d^3\mathbf{k}}{(2\pi)^4} \frac{e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)}}{k^2 - \omega^2/c^2} \tag{6.2.6}$$

We have an issue, there are two poles at $\omega = \pm ck$ in our integrand that must be integrated over. To simplify matters let us move to polar coordinates by setting the k_z-axis to point along $\mathbf{r}$. One then finds that $\mathbf{k} \cdot \mathbf{r} = kr\cos\theta$ and thus

$$G(\mathbf{r}, t) = -\frac{1}{(2\pi)^3} \int_0^\infty dk\, k^2 \int_{-\infty}^\infty d\omega \frac{e^{-i\omega t}}{k^2 - \omega^2/c^2} \int_0^\pi d\theta\, \sin\theta e^{ikr\cos\theta} \tag{6.2.7}$$

The integral in $d\theta$ can be evaluated by a simple substitution:

$$\int_0^\pi d\theta\, \sin\theta e^{ikr\cos\theta - \omega t} = -\frac{1}{ikr}\left[e^{ikr\cos\theta}\right]_0^\pi = 2\frac{\sin(kr)}{kr} \tag{6.2.8}$$

giving:

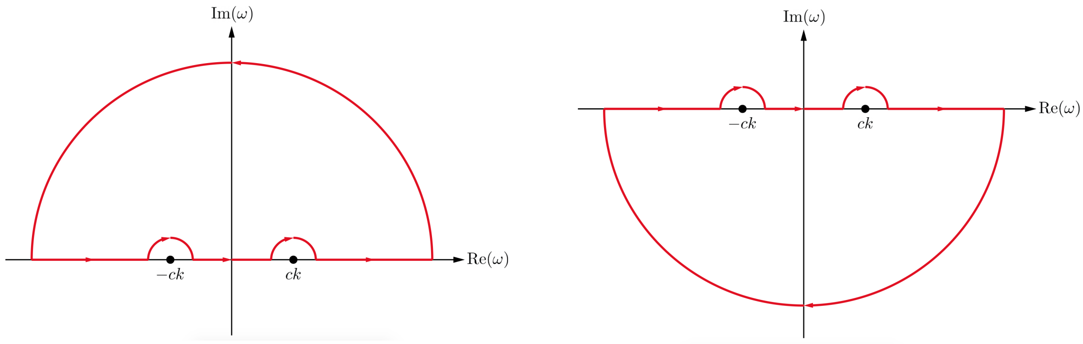
$$G(\mathbf{r}, t) = \frac{1}{4\pi^3} \int_0^\infty dk\, c^2 k^2 \frac{\sin(kr)}{kr} \int_{-\infty}^\infty d\omega \frac{e^{-i\omega t}}{(\omega - ck)(\omega + ck)} \tag{6.2.9}$$

(note the sign change due to the denominator). We can evaluate this integral in the complex ω-plane by choosing a contour running over $\mathrm{Re}(\omega)$ but jumping over the poles at $\omega = \pm ck$. There are several choices for such a contour, we present two that give the retarded and advanced Green's functions found earlier.

[1] note that depending in the $(-+++)$ metric $\mathsf{K} \cdot \mathsf{X} = \mathbf{k} \cdot \mathbf{r} - \omega t$.

Retarded Green's function

Suppose that $t < 0$ so that $e^{-i\omega t} \to 0$ as $\omega \to i\infty$. This suggests that we close our contour in the upper half plane ensuring that the integral due to the upper semi-circle does not give any contribution. This contour does not enclose either pole so by the residue theorem the integral vanishes.



(a) (b)

Figure 6.1. Contour for (a) $G_{ret}(t < 0)$ and (b) $G_{ret}(t > 0)$

Now suppose that $t > 0$. Then $e^{-i\omega t} \to 0$ as $\omega \to -i\infty$. This suggests that we close our contour in the lower half plane ensuring that the integral due t the lower semi-circle doe snot give any contribution. This time, we enclose both poles, so by the residue theorem:

$$\int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega t}}{(\omega - ck)(\omega + ck)} = -2\pi i \left(\frac{e^{-ickt}}{2ck} - \frac{e^{ickt}}{2ck} \right) = -\frac{2\pi}{ck} \sin(kct) \tag{6.2.10}$$

where the negative sign comes from the fact that the contour runs clockwise. Finally, we find that:

$$\begin{aligned} G_{ret}(\mathbf{r}, t) &= -\frac{1}{2\pi^2} \frac{c}{r} \int_0^\infty dk \sin(kr) \sin(kct) && (6.2.11) \\ &= \frac{1}{4\pi^2} \frac{c}{r} \frac{1}{4} \int_{-\infty}^{\infty} dk (e^{ikr} - e^{-ikr})(e^{ikct} - e^{-ikct}) dk && (6.2.12) \\ &= \frac{1}{4\pi^2} \frac{c}{r} \frac{1}{4} 2\pi (2\delta(r + ct) - 2\delta(r - ct)) && (6.2.13) \end{aligned}$$

Physically $r > 0 > -ct$ so $\delta(r + ct)$ can be safely neglected, giving:

$$G_{ret}(\mathbf{r}, t) = -\frac{1}{4\pi r} \delta(t_{ret}), \; t > 0 \tag{6.2.14}$$

where we used the identity $\delta(x/a) = |a|\delta(x)$. We rewrite this in the more usual notation:

$$\boxed{G_{ret}(\mathbf{x}, t, \mathbf{x}', t') = -\frac{1}{4\pi|\mathbf{x} - \mathbf{x}'|} \delta(t_{ret} - t')\Theta(t - t')} \tag{6.2.15}$$

Integrating the wave equation with this Green's function, we get that:

$$A_\mu = -\mu_0 \int d^3\mathbf{x}'\ G_{ret}(\mathbf{x}, t, \mathbf{x}', t') J_\mu(\mathbf{x}', t') \tag{6.2.16}$$

$$= \frac{\mu_0}{4\pi} \int d^3\mathbf{x}'\ \frac{J_\mu(\mathbf{x}', t_{ret})}{|\mathbf{x} - \mathbf{x}'|} \tag{6.2.17}$$

as found previously! It is easy to check that this potential satisfies the Lorentz gauge condition. Indeed:

$$\begin{aligned}
\partial^\mu A_\mu &= -\frac{\mu_0}{4\pi} \int d^3\mathbf{x}' \partial^\mu (G_{ret}(\mathbf{x}, t, \mathbf{x}', t')) J_\mu(\mathbf{x}', \mathbf{t}') \\
&= +\frac{\mu_0}{4\pi} \int d^3\mathbf{x}' \partial'^\mu (G_{ret}(\mathbf{x}, t, \mathbf{x}', t')) J_\mu(\mathbf{x}', \mathbf{t}') \\
&= +\frac{\mu_0}{4\pi} \int d^3\mathbf{x}' \partial'^\mu (G_{ret}(\mathbf{x}, t, \mathbf{x}', t') J_\mu(\mathbf{x}', \mathbf{t}')) \\
&\qquad - \frac{\mu_0}{4\pi} \int d^3\mathbf{x}' G_{ret}(\mathbf{x}, t, \mathbf{x}', t') \partial'^\mu (J_\mu(\mathbf{x}', \mathbf{t}'))
\end{aligned}$$

Taking the integral to infinity then the first vanishes by the divergence theorem (assuming localized sources), while the second vanishes due to charge conservation. Thus $\partial^\mu A_\mu = 0$ as desired.

Advanced potentials

With advanced potentials, we decide to integrate by skipping under the poles: The calculation is exactly similar, and gives:

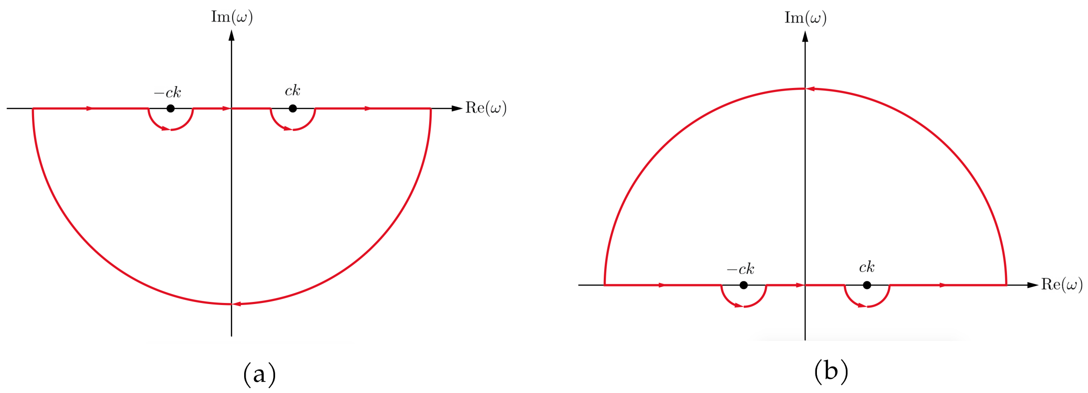


Figure 6.2. Contour for (a) $G_{ret}(t < 0)$ and (b) $G_{ret}(t > 0)$

$$\boxed{G_{adv}(\mathbf{x}, t, \mathbf{x}', t') = -\frac{1}{4\pi|\mathbf{x} - \mathbf{x}'|} \delta(t_{adv} - t') \Theta(t' - t)} \tag{6.2.18}$$

where

$$t_{adv} = t + \frac{|\mathbf{x} - \mathbf{x}'|}{c} \tag{6.2.19}$$

This gives the rather unphysical solution:

$$A_\mu(\mathbf{x},t) = \frac{\mu_0}{4\pi}\int d^3\mathbf{x}'\,\frac{J_\mu(\mathbf{x}',t_{adv})}{|\mathbf{x}-\mathbf{x}'|} \tag{6.2.20}$$

In QFT we will use a mix of these two propagators, the Feynman propagator, which can be found by using contours that go over one pole but under the other.

6.3 Jefimenko's equations

Now that we have found the retarded potentials:

$$\mathbf{A}(\mathbf{x}',t) = \frac{\mu_0}{4\pi}\int d^3\mathbf{x}'\,\frac{\mathbf{J}(\mathbf{x},t_{ret})}{|\mathbf{x}-\mathbf{x}'|}, \quad \phi(\mathbf{x},t) = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{x}'\,\frac{\rho(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|} \tag{6.3.1}$$

let us find the electromagnetic fields associated to them. Firstly, we find that:

$$\nabla\phi = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{x}'\left[\frac{\nabla\rho(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|} - \frac{\rho(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|^2}\nabla(|\mathbf{x}-\mathbf{x}'|)\right] \tag{6.3.2}$$

Using the chain rule:

$$\nabla\rho(\mathbf{x}',t_{ret}) = \frac{\partial\rho(\mathbf{x}',t_{ret})}{\partial t_{ret}}\nabla t_{ret} = \dot\rho(\mathbf{x}',t_{ret})\Big(-\frac{1}{c}\nabla(|\mathbf{x}-\mathbf{x}'|)\Big) \tag{6.3.3}$$

since $\frac{\partial}{\partial t_{ret}} = \frac{\partial}{\partial t}$. Consequently, using $\nabla(|\mathbf{x}-\mathbf{x}'|) = \frac{\mathbf{x}-\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|}$, one finds that

$$\nabla\phi = -\frac{1}{4\pi\epsilon_0}\int d^3\mathbf{x}'\left[\frac{\dot\rho(\mathbf{x}',t_{ret})}{c|\mathbf{x}-\mathbf{x}'|^2} + \frac{\rho(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|^3}\right](\mathbf{x}-\mathbf{x}') \tag{6.3.4}$$

We also find that:

$$\frac{\partial\mathbf{A}(\mathbf{x},t)}{\partial t} = \frac{1}{4\pi\epsilon_0}\frac{1}{c^2}\int d^3\mathbf{x}'\,\frac{\dot{\mathbf{J}}(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|} \tag{6.3.5}$$

yielding:

$$\boxed{\mathbf{E}(\mathbf{x},t) = \frac{1}{4\pi\epsilon_0}\int d^3\mathbf{x}'\left[\left(\frac{\dot\rho(\mathbf{x}',t_{ret})}{c|\mathbf{x}-\mathbf{x}'|^2} + \frac{\rho(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|^3}\right)(\mathbf{x}-\mathbf{x}') - \frac{\dot{\mathbf{J}}(\mathbf{x}',t_{ret})}{c^2|\mathbf{x}-\mathbf{x}'|}\right]} \tag{6.3.6}$$

Similarly, we find that:

$$\nabla\times\mathbf{A} = \frac{\mu_0}{4\pi}\int d^3\mathbf{x}\left[\frac{\nabla\times\mathbf{J}(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|} - \frac{\nabla(|\mathbf{x}-\mathbf{x}'|)}{|\mathbf{x}-\mathbf{x}'|^2}\cdot\mathbf{J}(\mathbf{x}',t_{ret})\right] \tag{6.3.7}$$

Now note that:

$$(\nabla\times\mathbf{J}(\mathbf{x}',t_{ret}))_i = \varepsilon_{ijk}\frac{\partial J^k}{\partial x_j} = \varepsilon_{ijk}\frac{\partial t_{ret}}{\partial x_j}\frac{\partial J^k}{\partial t_{ret}} \tag{6.3.8}$$

$$= -\frac{1}{c}\epsilon_{ijk}\frac{\partial|\mathbf{x}-\mathbf{x}'|}{\partial x_j}\frac{\partial J^k}{\partial t} = \left(\frac{1}{c}\dot{\mathbf{J}}\times\frac{\mathbf{x}-\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|}\right)_i \tag{6.3.9}$$

so we find that:

$$\mathbf{B}(\mathbf{x},t) = \frac{\mu_0}{4\pi}\int d^3\mathbf{x}'\left[\frac{\dot{\mathbf{J}}(\mathbf{x}',t_{ret})}{c|\mathbf{x}-\mathbf{x}'|^2} + \frac{\mathbf{J}(\mathbf{x},t_{ret})}{|\mathbf{x}-\mathbf{x}'|^3}\right]\times(\mathbf{x}-\mathbf{x}') \tag{6.3.10}$$

Suppose the sources are slowly varying, and can thus be Taylor expanded.

It is then easy to see that Jefimenko's equations reduce to the Coulomb and Biot-Savart laws:

Interestingly, the quasistatic approximation, which we took to be a zeroth order approximation, is actually a correct to first order due to this cancellation. Relativistic effects are thus only noticeable from second order corrections upwards.

6.4 Electric dipole radiation

Suppose we have a localized 4-current distribution $J_\mu(\mathbf{x},t)$ in a region $\mathcal{V}$. We could use the Jefimenko equations to compute the associated fields, but it is much simpler to compute the potentials and then differentiate them. The retarded potential reads:

$$A_\mu(\mathbf{x},t) = \frac{\mu_0}{4\pi}\int_{\mathcal{V}} d^3\mathbf{x}'\,\frac{J_\mu(\mathbf{x}',t_{ret})}{|\mathbf{x}-\mathbf{x}'|} \tag{6.4.1}$$

We now let $r = |\mathbf{x}|$ rather than $|\mathbf{x}-\mathbf{x}'|$. If $|\mathbf{x}-\mathbf{x}'| \gg d$ where d is the size of $\mathcal{V}$ then for all $\mathbf{x}' \in \mathcal{V}$ we may use the Taylor expansions:

$$|\mathbf{x}-\mathbf{x}'| \approx r - \frac{\mathbf{x}\cdot\mathbf{x}'}{r},\quad \frac{1}{|\mathbf{x}-\mathbf{x}'|} \approx \frac{1}{r^2} - \frac{\mathbf{x}\cdot\mathbf{x}'}{r^3} \tag{6.4.2}$$

We will also assume that the chracteristic time scale τ of the charges and currents is much larger than $\frac{d}{c}$. In other words, the charges can't change significantly over the time it takes for light to traverse $\mathcal{V}$. This allows us to Taylor expand the 4-current in $\frac{\mathbf{x}\cdot\mathbf{x}'}{rc}$:

$$J_\mu(\mathbf{x}',t-r/c+\mathbf{x}\cdot\mathbf{x}'/rc) \approx J_\mu(\mathbf{x}',t-r/c) + \dot{J}_\mu(\mathbf{x}',t-r/c)\frac{\mathbf{x}\cdot\mathbf{x}'}{rc} \tag{6.4.3}$$

Keeping only the first term gives the **dipole approximation**:

$$A_\mu \approx \frac{\mu_0}{4\pi r}\int_{\mathcal{V}} d^3\mathbf{x}'\ J_\mu(\mathbf{x}',t-r/c) \tag{6.4.4}$$

In the Electromagnetism volume we encountered the useful identity

$$\int d^3\mathbf{x}'\ \mathbf{J}(\mathbf{x}) = \dot{p} \tag{6.4.5}$$

To prove this, consider the continuity equation in component form:

$$\partial_i' J^i + \dot{\rho} = 0 \tag{6.4.6}$$

Integrating over $\mathbb{R}^3$ we find that:

$$\int d^3\mathbf{x}'\partial_i' J^i = -\int d^3\mathbf{x}'\dot{\rho} \implies \int d^3\mathbf{x}'\ x_j'\partial_i' J^i = -\int d^3\mathbf{x}'\ x_j'\dot{\rho} \tag{6.4.7}$$

$$\implies \int d^3\mathbf{x}'\ \partial_i(J^i x_j') = -\int d^3\mathbf{x}'\ x_j'\dot{\rho} + \int d^3\mathbf{x}\frac{\partial x_j'}{\partial x'^i}J^i \tag{6.4.8}$$

$$\implies \int d^3\mathbf{x}'\ \nabla\cdot(\mathbf{J}\otimes\mathbf{x}') = \int d^3\mathbf{x}'\left(\mathbf{J}-\dot{\rho}\mathbf{x}'\right) \tag{6.4.9}$$

Using Stokes' theorem for differential forms the integral on the LHS vanishes for a localized charge distribution that falls off at least as $\frac{1}{r}$, giving the desired result. Applying this to (6.4.4) gives:

$$\boxed{\mathbf{A}(r,t) \approx \frac{\mu_0}{4\pi r}\dot{\mathbf{p}}(t-r/c)} \tag{6.4.10}$$

which is indeed a dipole! The magnetic field is then found to be:

$$\mathbf{B} \approx \frac{\mu_0}{4\pi}\left(\frac{1}{r^2}(\nabla r)\dot{\mathbf{p}}(t-r/c) + \frac{1}{r}(\nabla t - r/c)\times\dot{\mathbf{p}}(t-r/c)\right) \tag{6.4.11}$$

$$= -\frac{\mu_0}{4\pi r^2}\hat{\mathbf{x}}\times\dot{\mathbf{p}}(t-r/c) - \frac{\mu_0}{4\pi rc}\hat{\mathbf{x}}\times\ddot{\mathbf{p}}(t-r/c) \tag{6.4.12}$$

Suppose the source oscillates at a frequency ω. Then $\ddot{\mathbf{p}} \sim \omega\dot{\mathbf{p}}$ so the first term is negligible as long as $r \gg \frac{c}{\omega}$, that is as long as we are in the **far-field limit**. We have therefore found that:

$$\boxed{\mathbf{B}(\mathbf{x},t) \approx -\frac{\mu_0}{4\pi rc}\hat{\mathbf{x}}\times\ddot{\mathbf{p}}(t-r/c)} \tag{6.4.13}$$

Let us now compute the scalar potential by using the Lorentz gauge condition:

$$\frac{\partial\phi}{\partial t} = -c^2\nabla\cdot\mathbf{A} \tag{6.4.14}$$

From (6.4.10) we get:

$$\nabla\cdot\mathbf{A} = \frac{\mu_0}{4\pi}\left(\frac{1}{r}\nabla\cdot\dot{\mathbf{p}}(t-r/c) - \frac{1}{r^2}(\nabla r)\cdot\dot{\mathbf{p}}(t-r/c)\right) \tag{6.4.15}$$

$$= \frac{\mu_0}{4\pi}\left(\frac{1}{r}\ddot{\mathbf{p}}(t-r/c)\cdot\nabla(t-r/c) - \frac{\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c)}{r^2}\right) \tag{6.4.16}$$

$$= -\frac{\mu_0}{4\pi}\left(\frac{\hat{\mathbf{x}}\cdot\ddot{\mathbf{p}}(t-r/c)}{cr} + \frac{\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c)}{r^2}\right) \tag{6.4.17}$$

so:

$$\frac{\partial\phi}{\partial t} = \frac{1}{4\pi\epsilon_0}\left(\frac{\hat{\mathbf{x}}\cdot\ddot{\mathbf{p}}(t-r/c)}{cr} + \frac{\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c)}{r^2}\right) \tag{6.4.18}$$

$$\implies \phi(r,t) = \frac{1}{4\pi\epsilon_0}\left(\frac{\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c)}{cr} + \frac{\hat{\mathbf{x}}\cdot\mathbf{p}(t-r/c)}{r^2}\right) \tag{6.4.19}$$

Again, in the far-field approximation $r \gg \frac{c}{\omega}$ so the first term dominates:

$$\boxed{\phi(r,t) \approx \frac{1}{4\pi\epsilon_0 rc}\hat{\mathbf{x}} \cdot \dot{\mathbf{p}}(t-r/c)} \tag{6.4.20}$$

Taking the gradient of the potential gives:

$$\begin{aligned}
\nabla\phi &= \frac{1}{4\pi\epsilon_0 c}\left[\frac{1}{r}\nabla(\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c)) - \frac{\hat{\mathbf{x}}}{r^2}\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c)\right] && (6.4.21)\\
&= \frac{1}{4\pi\epsilon_0 c}\left[\frac{1}{r}\Big((\nabla\cdot\hat{\mathbf{x}})\dot{\mathbf{p}}(t-r/c)) + (\nabla\cdot\dot{\mathbf{p}}(t-r/c))\hat{\mathbf{x}}\Big) - \frac{1}{r^2}(\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c))\hat{\mathbf{x}}\right] && (6.4.22)\\
&= \frac{1}{4\pi\epsilon_0 c}\left[\frac{2}{r^2}\dot{\mathbf{p}}(t-r/c) - \frac{1}{rc}(\ddot{\mathbf{p}}(t-r/c)\cdot\hat{\mathbf{x}})\hat{\mathbf{x}} - \frac{1}{r^2}(\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c))\hat{\mathbf{x}}\right] && (6.4.23)\\
&\approx -\frac{1}{4\pi\epsilon_0 rc}(\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c))\hat{\mathbf{x}} && (6.4.24)
\end{aligned}$$

so that:

$$\mathbf{E}(\mathbf{x},t) \approx \frac{1}{4\pi\epsilon_0 rc}(\hat{\mathbf{x}}\cdot\dot{\mathbf{p}}(t-r/c))\hat{\mathbf{x}} - \frac{1}{4\pi\epsilon_0 rc}\dot{\mathbf{p}}(t-r/c) \tag{6.4.25}$$

or equivalently:

$$\boxed{\mathbf{E}(\mathbf{x},t) = \frac{1}{4\pi\epsilon_0 rc}\hat{\mathbf{x}}\times(\hat{\mathbf{x}}\times\ddot{\mathbf{p}}(t-r/c)} \tag{6.4.26}$$

6.5 Dipole radiation power

6.6 Magnetic dipole radiation

6.7 Lienard-Wiechart potentials

Suppose we have a point particle with charge distribution $\rho(\mathbf{x},t) = q\delta^3(\mathbf{x}-\mathbf{r}(t))$ where $\mathbf{r}(t)$ is the position of the particle at time t. The scalar potential reads:

$$\phi(\mathbf{x},t) = \frac{q}{4\pi\varepsilon_0}\int d^3\mathbf{x}\frac{\delta^3(\mathbf{x}'-\mathbf{r}(t_{ret}))}{|\mathbf{x}-\mathbf{x}'|} \tag{6.7.1}$$

This integral does not give the usual time-independent potential because t_{ret} depends on $\mathbf{x}'$ too. We fix this as follows, first we add an integration over t' (note that t' does not mean anything, it is a dummy variable):

$$\begin{aligned}
\phi(\mathbf{x},t) &= \frac{q}{4\pi\varepsilon_0}\int dt'\int d^3\mathbf{x}\frac{\delta^3(\mathbf{x}'-\mathbf{r}(t'))}{|\mathbf{x}-\mathbf{x}'|}\delta(t'-t_{ret}) && (6.7.2)\\
&= \frac{q}{4\pi\epsilon_0}\int dt'\,\frac{1}{|\mathbf{x}-\mathbf{r}(t')}\delta(t-t'-|\mathbf{x}-\mathbf{r}(t')|/c) && (6.7.3)
\end{aligned}$$

Now let $\mathbf{R}(t) = \mathbf{x} - \mathbf{r}(t)$ and $f(t') = t' + |\mathbf{R}(t)|/c$. We find that:

$$\phi(\mathbf{x},t) = \frac{q}{4\pi\epsilon_0}\int dt' \frac{1}{R(t')}\delta(t - f(t')) = \frac{q}{4\pi\epsilon_0}\int df \frac{dt'}{df}\frac{1}{|\mathbf{R}(t')|}\delta(t - f(t')) \tag{6.7.4}$$

$$= \frac{q}{4\pi\epsilon_0}\left[\frac{dt'}{df}\frac{1}{|\mathbf{R}(t')|}\right]_{f(t')=t} \tag{6.7.5}$$

We quickly find that:

$$\frac{df}{dt'} = 1 + \frac{1}{c}\frac{d|\mathbf{R}(t)|}{dt} = 1 - \frac{\mathbf{v}(t')\cdot\mathbf{R}(t')}{|\mathbf{R}(t')|} \tag{6.7.6}$$

where we defined $\mathbf{v}(t) = \dot{\mathbf{r}}(t)$ to be the particle velocity. Consequently:

$$\phi(\mathbf{x},t) = \frac{q}{4\pi\epsilon_0}\left[\frac{c}{c|\mathbf{R}(t')| - \mathbf{R}(t')\cdot\mathbf{v}(t')}\right]_{f(t')=t} \tag{6.7.7}$$

Note that this expression must be evaluated at t' such that $f(t') = t \implies t' = t - \frac{|\mathbf{x}-\mathbf{r}(t')|}{c}$. Similarly one finds that:

$$\mathbf{A}(\mathbf{x},t) = \frac{q}{4\pi\epsilon_0 c}\left[\frac{c\mathbf{v}(t')}{c|\mathbf{R}(t')| - \mathbf{R}(t')\cdot\mathbf{v}(t')}\right]_{f(t')=t} \tag{6.7.8}$$

Finally, (6.7.7) and (6.7.8) can be summarized into a 4-vector equation:

$$\boxed{A_\mu(\mathbf{x},t) = -\frac{q}{4\pi\epsilon_0 c}\frac{U_\mu(t')}{R^\nu(t')U_\nu(t')}} \tag{6.7.9}$$

where $R^\nu(t') = (|\mathbf{R}(t')|, \mathbf{R}(t'))$.

7 Spinors

The Principle of Equivalence

8

The Einstein Field Equations

9

10 Swarzchild's solution and Black holes

Part II

Quantum Field Theory

Classical field theory

11

11.1 Why fields?

There are two approaches to quantum field theory. In one approach, the particles are regarded as fundamental giving rise to fields e.g. photons give rise to the EM field. The other viewpoint is that the fields are fundamental, and they give rise to particles when quantized i.e. EM field quantization gives rise to photons.

One reason we should think in terms of fields is locality: a perturbation has a local influence and does not propagate instantaneously, and fields naturally behave like this.

Also, all bosons (and fermions) are indistinguishable. Take an electron from the edge of the universe and compare it to an electron in a coffee cup and they will have the exact same properties, almost as if there was no “error in their production process”. This can be explained by regarding any two electrons as both belong to the same field, so of course they must be identical.

Furthermore,the total particle number is not conserved in relativistic quantum effects. In a typical high energy collision (inelastic), two particles can give rise to several other particles of different nature. Consequently, one cannot take the Schrodinger equation (or any single-particle framework) and “relativize” it without dealing with problems such as negative probabilities and unbounded energy levels, all due to the loss of particle number conservation. The fix is, once again, fields.

As an illustrative example, consider a particle of mass m in a box of size L. By Heisenberg's relation, $\Delta p \gtrapprox \frac{\hbar}{L}$, and thus in some frame we will have that $\Delta E \gtrapprox \frac{\hbar c}{L}$. However, if $\Delta E \gtrapprox 2mc^2$ then it is possible to create particle-antiparticle pairs out of the vacuum, thus violating the conservation of particle number. This occurs when $L \lessapprox \frac{\hbar}{2mc}$ where $\lambda = \frac{\hbar}{mc}$ is known as the **Compton wavelength**. Just like the de Broglie wavelength delineates the limit where a particle starts to exhibit wave-like properties, the Compton wavelength delineates the limit where it no longer makes sense to talk about particles.

Finally, recall that in undergraduate quantum mechanics we took classical observables and quantized them by promoting them to quantum operators. Similarly, in quantum field theory we will take classical fields and quantize them by promoting them to quantum fields. However, to do so we must first get comfortable with manipulating classical fields.

Units

QFT is one of those subjects where we can afford to treat units more or less as we wish. More specifically, we will be working in **natural units** where $\hbar = c = 1$, allowing us to express all quantities in terms of mass/energy.

11.2 What is a field?

A field is a map that assigns a quantity at every point in space and time. It follows that while in classical mechanics we have a finite number of degrees of freedom

$$(q_1(t), ..., q_n(t), p_1(t), ..., p_m(t))$$

In field theory, one the other hand, we have an infinite number of degrees of freedom $\phi_\mu(t)$ corresponding to the continuum nature of space and time. For example, the electric field $\mathbf{E}(\mathbf{x}, t)$ and magnetic field $\mathbf{B}(\mathbf{x}, t)$ are, as the name suggests, fields. More precisely, they are vector fields in $\mathbb{R}^3$.

The evolution of a field is given by a Lagrangian $L(\phi, \dot{\phi}, \nabla\phi)$. We define the **Lagrangian density** $\mathcal{L}(\phi_a, \partial_\mu\phi_a)$ to satisfy:

$$L(t) = \int d^3\mathbf{x}\ \mathcal{L}(\phi_a, \partial_\mu\phi_a) \tag{11.2.1}$$

so that the action reads:

$$S = \int dt\ L(t) = \int d^4\mathsf{x}\ \mathcal{L}(\phi_a, \partial_\mu\phi_a) \tag{11.2.2}$$

Note that since we are treating space and time on equal footing, we shall not consider lagrangians with $\nabla\phi, \nabla^2\phi$ and higher order spatial derivatives. On the other hand, in condensed matter field theory where relativistic effects are negligible, we are allowed to consider lagrangians with such terms.

The equations of motion for fields can be derived by the principle of least action:

> **Principle of least action:** if we fix the value of the field on some boundary and vary the field, the variation in the action will be zero.

Consequently:

$$\delta S = \int d^4\mathsf{x} \left[\frac{\partial\mathcal{L}}{\partial\phi_a}\delta\phi_a + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\delta(\partial_\mu\phi_a)\right] \tag{11.2.3}$$

$$= \int d^4\mathsf{x} \left[\frac{\partial\mathcal{L}}{\partial\phi_a} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right)\right]\delta\phi_a + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_\mu)}\delta\phi_a\right) \tag{11.2.4}$$

The boundary term vanishes for any infinitesimal field variation $\delta\phi_a(\mathbf{x}, t)$ as long as it decays at $x \to \infty$ and $\delta\phi_a(\mathbf{x}, t_i) = \phi_a(\mathbf{x}, t_i) = 0$. Requiring (11.2.3) to vanish identically for all $\delta\phi_a$ gives the **Euler-Lagrange field equations** (ELF):

$$\boxed{\frac{\partial \mathcal{L}}{\partial \phi_a} - \partial_\mu\left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)}\right) = 0} \tag{11.2.5}$$

11.3 Lorentz invariance

Since we are interested in unifying special relativity with quantum mechanics, we should define what a **Lorentz invariant field** is. Suppose we have a field $\phi(\mathsf{x})$ which solves the ELF equations, and suppose we perform an active Lorentz transformation [1]

$$\phi(\mathsf{x}) \mapsto \phi'(\mathsf{x}) \equiv \phi(\Lambda^{-1}\mathsf{x}) \tag{11.3.1}$$

To see why this should hold for active transformations, consider a 2D scalar field. If I rotate this field by some angle, then the new value of the field at some point should be equal to the value of the field at the original, unrotated point. For a vector field, not only do we have to rotate the coordinates, we should also do this for the direction of the field:

$$A_\mu(x) \mapsto A'_\mu(x) \equiv \Lambda_\mu{}^\nu A_\nu(\Lambda^{-1}x) \tag{11.3.2}$$

For a theory to be Lorentz invariant we need $\phi'(x) = \phi(\Lambda^{-1}x)$ to be a solution too. This can be ensured by checking that the action is Lorentz invariant.

For example, consider the action

$$S[\phi] = \int d^3x \left(\frac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - \frac{1}{2}m^2\phi^2\right) \tag{11.3.3}$$

We claim that this action is invariant under Lorentz transformations i.e. $S[\phi] = S[\phi']$. Indeed we have that the differential stransform to

$$\frac{1}{2}\eta^{\mu\nu}\partial_\mu\phi(\lambda^{-1}x)\partial_\nu\phi(\lambda^{-1}x) = \frac{1}{2}\eta^{\mu\nu}\partial_\mu y^\alpha\partial'_\alpha\phi(y)\partial_\nu y^\beta\partial'_\beta\phi(y) \tag{11.3.4}$$

where we let $y = \Lambda^{-1}x$ and $\partial'_\mu = \frac{\partial}{\partial y^\mu}$. Consequently we have that

$$S[\phi'] = \int d^3x\left[\frac{1}{2}\eta^{\mu\nu}(\Lambda^{-1})^\alpha{}_\mu(\Lambda^{-1})^\beta{}_\nu\partial'{}_\alpha\phi(y)\partial'{}_\beta\phi(y) - \frac{1}{2}m^2\phi^2(y)\right] \tag{11.3.5}$$

We now use the defining property of the Lorentz group

$$\Lambda^\alpha{}_\mu\eta^{\mu\nu}\Lambda^\beta{}_\nu = \eta^{\alpha\beta} \tag{11.3.6}$$

so that

$$S[\phi'] = \int d^3x\left[\frac{1}{2}\eta^{\alpha\beta}\partial'{}_\alpha\phi(y)\partial'{}_\beta\phi(y) - \frac{1}{2}m^2\phi^2(y)\right] = S[\phi'] \tag{11.3.7}$$

as desired.

[1] we boost the field rather than performing a passive transformation and boosting the coordinates $\phi(\mathsf{x}) \mapsto \phi'(\mathsf{x}) \equiv \phi(\Lambda\mathsf{x})$

11.4 Symmetries and Noether's Theorem

We define a symmetry of an action $S[\phi]$ as a transformation which can be performed on any field ϕ such that $\delta S = 0$.

Noether's Theorem: every continuous symmetry of the action S gives rise to a conserved current $j^\mu(x)$ such that: $\partial_\mu j^\mu = 0$

Proof. Indeed, consider the infinitesimal transformation (which we are allowed to consider for continuous symmetries):

$$\phi_a(x) \mapsto \phi_a(x) + \delta\phi_a(x) \tag{11.4.1}$$

For this transformation to be a symmetry of the action, we need $\delta\mathcal{L}$ to change at most by a full differential $\delta\mathcal{L} = \partial_\mu F^\mu$ which vanishes when integrated to get the action. For an arbitrary transformation of the field we then find that:

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi_a}\delta\phi_a + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\delta(\partial_\mu\phi_a) \tag{11.4.2}$$

$$= \left(\frac{\partial\mathcal{L}}{\partial\phi_a} - \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right)\delta\phi_a + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\delta\phi_a\right) \tag{11.4.3}$$

but the first term must vanish by the ELF equations. Hence, for this to be a symmetry transformation then we must require that:

$$\delta\mathcal{L} = \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\delta\phi_a\right) = \partial_\mu F^\mu(\phi) \tag{11.4.4}$$

implying that:

$$j^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\delta\phi_a - F^\mu(\phi) \tag{11.4.5}$$

is conserved. ■

On-shell vs. Off-shell

At a first glance, it seems like the principle of least action ensures that any transformation is a symmetry of an action S, giving an uncountably infinite number of symmetries!

However, there is a difference between the definition of symmetry and the principle of least action. A symmetry transformation $\phi \mapsto \phi + \delta\phi$ is a symmetry if $\delta S = 0$ for any ϕ regardless of whether it minimizes the action or not. A solution ϕ to the ELF instead satisfies $\delta S = 0$ for all possible $\phi \mapsto \phi + \delta\phi$. Noether's theorem then states that given a symmetry transformation of the action, when applied to a solution to the ELF equation this symmetry will produce a conserved current (which is why we could use the ELF equations in our proof of Noether's theorem).

Statements that are made on fields that minimize the action will often be referred to as **on-shell**, while statements on all possible fields are **off-shell**. Thus the definition of a symmetry transformation is off-shell, while Noether's theorem states that the current is conserved on-shell.

Conserved charges

Given a conserved current, we must have an associated **conserved charge**:

$$Q = \int_{\mathbb{R}^3} d^3\mathbf{x}\, j^0 \tag{11.4.6}$$

since:

$$\frac{dQ}{dt} = \int_{\mathbb{R}^3} d^3\mathbf{x} \frac{\partial j^0}{\partial t} = -\int_{\mathbb{R}^3} d^3\mathbf{x} \nabla \cdot \mathbf{j} = 0 \tag{11.4.7}$$

for bounded currents at infinity. Note also that given a finite volume $\mathcal{V}$ then:

$$\frac{dQ_{\mathcal{V}}}{dt} = \int_{\mathcal{V}} d^3\mathbf{x} \frac{\partial j^0}{\partial t} = -\int_{\partial\mathcal{V}} d\mathbf{S} \cdot \mathbf{j} \tag{11.4.8}$$

so not only is charge conserved **globally**, it is also **locally conserved**. In simpler terms: if charge gets smaller in some volume then there must be a current flux out of this region's to compensate.

Consider an infinitesimal translation $x^\mu \mapsto x^\mu + \epsilon^\mu$ so that the field ϕ_a and the Lagrangian $\mathcal{L}(\phi_a)$ acting on it:

$$\phi_a(x) \mapsto \phi_a(x) - \epsilon^\mu \partial_\mu \phi_a(x),\ \mathcal{L}(x) \mapsto \mathcal{L}(x) + \epsilon^\mu \partial_\mu \mathcal{L}(x) \tag{11.4.9}$$

where we assume that the Lagrangian has no explicit x dependence. Since the Lagrangian changes by a full differential, our action is **translationally invariant** giving rise to 4 conserved currents (one for each possible translation in Minkowski space):

$$(j^\mu)_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} \partial_\nu \phi_a - \delta^\mu_\nu \mathcal{L} \tag{11.4.10}$$

This current is known as the **Stress-energy tensor**. The corresponding conserved quantities are:

$$E = \int d^3\mathbf{x}\, T^{00} \ \text{ which is the total field energy} \tag{11.4.11}$$

$$p^i = \int d^3\mathbf{x}\, T^{0i} \ \text{ which is the total field momentum} \tag{11.4.12}$$

Again considering the following field:

$$\mathcal{L} = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 \tag{11.4.13}$$

then we see that:

$$E =,\ p^i = \tag{11.4.14}$$

11.5 Klein-Gordon field

Consider the following Lagrangian density for a set of three scalar real fields ϕ_a, $a = 1, 2, 3$:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi_a\partial^\mu\phi_a - \frac{1}{2}m^2\phi_a\phi_a \tag{11.5.1}$$

This lagrangian is invariant under SO(3) rotations. Indeed let us consider an infinitesimal rotation by an angle θ about the axis $\hat{\mathbf{n}}$:

$$\mathsf{R}_{\mathbf{n}}(\theta)\phi_a = \phi_a + \theta\epsilon_{abc}n_b\phi_c \tag{11.5.2}$$

The lagrangian after this rotation is given by (we can use the same a, b, c indices as all second order terms will be negligible):

$$\begin{aligned}\mathcal{L}' = &\frac{1}{2}\partial_\mu(\phi_a + \theta\epsilon_{abc}n_b\phi_c)\partial^\mu(\phi_a + \theta\epsilon_{abc}n_b\phi_c)\\ &\qquad - \frac{1}{2}m^2(\phi_a + \theta\epsilon_{abc}n_b\phi_c)(\phi_a + \theta\epsilon_{abc}n_b\phi_c)\\ = &\,\mathcal{L} + \frac{1}{2}\theta\epsilon_{abc}n_b\left[(\partial_\mu\phi_c\partial^\mu\phi_a + \partial^\mu\phi_c\partial_\mu\phi_a) - 2m^2\phi_a\phi_c\right] + o(\theta^2)\end{aligned}$$

Now note that:

$$\epsilon_{abc}(\partial_\mu\phi_c\partial^\mu\phi_a + \partial^\mu\phi_c\partial_\mu\phi_a) = \epsilon_{abc}(\partial_\mu\phi_c\partial^\mu\phi_a - \partial^\mu\phi_a\partial_\mu\phi_c) = 0 \tag{11.5.3}$$

and recall that $\boldsymbol{\phi} \cdot (\mathbf{n} \times \boldsymbol{\phi}) = 0 \implies \epsilon_{abc}n_b\phi_c = 0$. Then we find that $\mathcal{L}' = \mathcal{L}$ so SO(3) is indeed a symmetry of this lagrangian.

The equations of motion are given by:

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right) = \frac{\partial\mathcal{L}}{\partial\phi_a} \tag{11.5.4}$$

where:

$$\frac{\partial\mathcal{L}}{\partial\phi_a} = -m^2\phi_a \tag{11.5.5}$$

and

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\right) = \partial_\mu\partial^\mu\phi_a = \Box^2\phi_a \tag{11.5.6}$$

so we obtain:

$$(\Box^2 + m^2)\phi_a = 0 \tag{11.5.7}$$

known as the Klein-Gordon equation. By Noether's theorem, there must be a conserved current associated to the SO(3) symmetry. It is given by:

$$J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi_a)}\delta\phi_a - F^\mu \tag{11.5.8}$$

but since $\partial_\mu F^\mu = \delta\mathcal{L} = 0$ we can set $F^\mu = 0$. Then we see that since $\delta\phi_a = \epsilon_{abc}n_b\phi_c$ the conserved current is:

$$J^\mu = \epsilon_{abc}(\partial^\mu \phi_a) n_b \phi_c \tag{11.5.9}$$

giving a conserved charge:

$$Q = \int d^3x J^0 = \int d^3x \epsilon_{abc} \dot{\phi}_a n_b \phi_c \tag{11.5.10}$$

Now we can without loss of generality align our axes so that **n** points along one of the 3-axes, hence $n_b = \delta^b_n$ where $n = 1, 2, 3$. Then we see that we have three individual conserved charges:

$$Q_n = \int d^3x \epsilon_{abc} \dot{\phi}_a \delta^b_n \phi_c = \int d^3x \epsilon_{anc} \dot{\phi}_a \phi_c = -\int d^3x \epsilon_{nac} \dot{\phi}_a \phi_c \tag{11.5.11}$$

We can also check that $\partial_\mu J^\mu$ using the Klein-Gordon equation:

$$\begin{aligned} \partial_\mu J^\mu &= \partial_\mu(\epsilon_{abc}(\partial^\mu \phi_a) n_b \phi_c) = \epsilon_{abc} n_b (\partial^\mu \phi_a \partial_\mu \phi_c + \phi_c \Box^2 \phi_a) && (11.5.12)\\ &= \epsilon_{abc} n_b (\partial^\mu \phi_a \partial_\mu \phi_c - m^2 \phi_a \phi_c) && (11.5.13)\\ &= 0 && (11.5.14) \end{aligned}$$

where we used the fact that $\epsilon_{abc}\partial^\mu \phi_a \partial_\mu \phi_c n_b = g^\mu_\mu \epsilon_{abc} \partial^\mu \phi_a \partial^\mu \phi_c n_b = 0$.

11.6 Global symmetries

A **global** or **internal** symmetry is a transformation that involves the fields only and acts homogeneously on space-time.

For example, consider the complex scalar field ϕ governed by the Lagrangian:

$$\mathcal{L} = \partial_\mu \phi^* \partial^\mu \phi - V|\phi|^2 \tag{11.6.1}$$

Consider the following transformation:

$$\phi \mapsto e^{i\alpha}\phi \implies \delta\phi = i\alpha\phi,\ \delta\phi^* = -i\alpha\phi^* \tag{11.6.2}$$

where to compute $\delta\phi$ we performed a taylor expansion to first order. This is clearly a symmetry, and it is easy to see that the associated conserved current is:

$$j^\mu = i(\partial^\mu \phi^*) - (\partial^\mu \phi)\phi^* \tag{11.6.3}$$

There is a nice trick that can be used to compute these conserved currents for global symmetries. Suppose we have found a global symmetry $\delta\phi = \alpha\phi$ where α is a constant. We now redo the transformation making $\alpha(x)$ depend on space-time. This is no longer a symmetry $\delta\mathcal{L} \neq 0$, but must become one as we make α constant. This can only happen if $\delta\mathcal{L}$ depends on the derivative of α so:

$$\delta\mathcal{L} = \partial_\mu \alpha(x) h^\mu \implies \delta S = -\int d^4x\ \alpha(x)\partial_\mu h^\mu \tag{11.6.4}$$

Note however that the action must be stationary so $\delta S = 0$, so the integrand must vanish identically, yielding:

$$\partial_\mu h^\mu = 0 \tag{11.6.5}$$

We may identify the conserved current as $j^\mu = h^\mu$, this is much quicker!

11.7 Electromagnetic field

Consider the following lagrangian:

$$\mathcal{L} = -\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} - \mu_0 A_\beta J^\beta \tag{11.7.1}$$

Its equation of motion is given by:

$$\frac{\partial \mathcal{L}}{\partial A_\nu} = -\mu_0 J^\nu, \ \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} = \partial_\mu \frac{\partial}{\partial(\partial_\mu A_\nu)} \Big(-\frac{1}{2} (\partial_\alpha A_\beta) F^{\alpha\beta} \Big) \tag{11.7.2}$$

$$= -\frac{1}{2} \partial_\mu \Big[F^{\mu\nu} + \partial_\alpha A_\beta \frac{\partial}{\partial(\partial_\mu A_\nu)} (\partial^\alpha A^\beta - \partial^\beta A^\alpha) \Big] \tag{11.7.3}$$

$$= -\frac{1}{2} \partial_\mu [F^{\mu\nu} + \partial_\alpha A_\beta (g^{\alpha\mu} g^{\beta\nu} - g^{\alpha\nu} g^{\beta\mu})] = -\partial_\mu F^{\mu\nu} \tag{11.7.4}$$

$$\implies \partial_\mu F^{\mu\nu} = \mu_0 J^\nu \tag{11.7.5}$$

which reproduces the inhomogeneous Maxwell equations. It follows that (11.7.1) must be the lagrangian for an electromagnetic field.

One important property of the lagrangian is that it is **not** gauge invariant, but transforms quite nicely under gauge transformations which leads to charge conservation. Indeed, consider a general gauge transformation:

$$A_\mu \to A_\mu + \partial_\nu \partial^\nu \chi \tag{11.7.6}$$

The electromagnetic field Lagrangian is gauge invariant, since

$$\mathcal{L} \to -\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} - \mu_0 (A_\beta + \tag{11.7.7}$$

11.8 The Hamiltonian formulation

The Lagrangian formulation is so powerful and useful in QFT because it is a manifestly covariant framework. On the other hand, we know from analytical mechanics that we have an equivalent Hamiltonian formulation.

We define the **momentum conjugate to** $\phi_a(x)$ as:

$$\pi(x) = \frac{\partial \mathcal{L}}{\partial \dot{\phi}_a} \tag{11.8.1}$$

and the **Hamiltonian density** as a Legendre transform of the Lagrangian with respect to

$\dot{\phi}_a$:

$$\mathcal{H} = \pi(x)\dot{\phi}(x) - \mathcal{L}(\phi, \partial_\mu \phi) \tag{11.8.2}$$

We see that the Hamiltonian density is no longer manifestly Lorentz covariant as it picks out a time derivative. Consider as an example Hamilton's equations:

$$\dot{\phi}(x) = \frac{\partial \mathcal{H}}{\partial \pi},\ \dot{\phi}(x) = -\frac{\partial \mathcal{H}}{\partial \phi} \tag{11.8.3}$$

The theory is still invariant, but it is not clear at first sight unlike the Lagrangian theory.

Canonical quantization

12

12.1 Quantizing scalar fields

Quantum fields

To quantize classical mechanics, we took the Darboux coordinates (q_a, p^a) satisfying the symplectic algebra:

$$\{q_a, p^b\} = 1, \ \{q_a, q_b\} = \{p^a, p^b\} = 0 \tag{12.1.1}$$

and promoted them to operators $\hat{q}_a, \hat{p}^b$ satisfying the Poisson algebra:

$$\{\hat{q}_a, \hat{p}^b\} = i\delta_a^b, \ \{\hat{q}_a, \hat{q}_b\} = \{\hat{p}^a, \hat{p}^b\} = 0 \tag{12.1.2}$$

Similarly, we can promote classical fields $\phi(\mathbf{x})$ and $\pi(\mathbf{x})$. We are working in the Schrodinger picture where the fields depend on space coordinates only and have no time-dependence. Furthermore we require these quantum fields to satisfy the commutation relations:

$$[\phi_a(\mathbf{x}, \phi_b(\mathbf{y})] = i\delta_a^b\delta^3(\mathbf{x}-\mathbf{y}), \tag{12.1.3a}$$
$$[\phi_a(\mathbf{x}, \phi_b(\mathbf{y}] = [\phi_a(\mathbf{x}), \phi_b(\mathbf{y})] = 0 \tag{12.1.3b}$$

As in typical QM, all information about our system lies in the spectrum of the Hamiltonian. This is unfortunately very hard for most quantum fields due to the infinite number of degrees of freedom. However, in **free field theories**, we can separate these degrees of freedom and integrate them separately. Free fields usually have Lagrangians that are quadratic in the fields giving linear equations of motion. We have already seen a classical free field theory, namely the Klein-Gordon field governed by the equation:

$$\partial_\mu\partial^\mu\phi + m^2\phi = 0 \tag{12.1.4}$$

Let us take the Fourier transform:

$$\phi(\mathbf{x}, t) = \int \frac{d^3\mathbf{p}}{(2\pi)^3} e^{i\mathbf{p}\cdot\mathbf{x}}\tilde{\phi}(\mathbf{p}, t) \tag{12.1.5}$$

and substitute into the KG equation:

$$\left(\frac{\partial^2}{\partial t^2} + \omega_\mathbf{p}^2\right)\tilde{\phi}(\mathbf{p}, t), \qquad \omega_\mathbf{p} = \sqrt{\mathbf{p}^2 + m^2} \tag{12.1.6}$$

We get a harmonic oscillator with frequency $\omega_{\mathbf{p}}$ for each momentum mode $\mathbf{p}$, so the coefficients of each plane wave mode in our ansatz will oscillate in time (this is expected as taking the FT of a dirac delta will give a sinusoid). Consequently we find that:

$$\phi(\mathbf{x},t) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1}{\sqrt{2\omega_{\mathbf{p}}}}(A^+_{\mathbf{p}}e^{i\omega_{\mathbf{p}}t} + A^-_{\mathbf{p}}e^{-i\omega_{\mathbf{p}}t})e^{i\mathbf{p}\cdot\mathbf{x}} \tag{12.1.7}$$

where the $\frac{1}{\sqrt{2\omega_{\mathbf{p}}}}$ factor is inserted by convention, and will make the transition to quantum fields more accessible. We can rewrite this as:

$$\phi(\mathbf{x},t) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1}{\sqrt{2\omega_{\mathbf{p}}}}(A^+_{-\mathbf{p}}e^{-i(\mathbf{p}\cdot\mathbf{x}-\omega_{\mathbf{p}}t)} + A^-_{\mathbf{p}}e^{i(\mathbf{p}\cdot\mathbf{x}-\omega_{\mathbf{p}})t}) \tag{12.1.8}$$

Now since ϕ must be a real scalar field, we require $A^+_{-\mathbf{p}} = (A^-_{\mathbf{p}})^*$. So, by setting $A^-_{\mathbf{p}} \equiv A_{\mathbf{p}}$ then we find that:

$$\boxed{\phi(\mathbf{x},t) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1}{\sqrt{2\omega_{\mathbf{p}}}}(A_{\mathbf{p}}e^{i(\mathbf{p}\cdot\mathbf{x}-\omega_{\mathbf{p}})t} + A^*_{\mathbf{p}}e^{-i(\mathbf{p}\cdot\mathbf{x}-\omega_{\mathbf{p}}t)})} \tag{12.1.9}$$

and similarly recalling that $\pi(\mathbf{x},t) = \dot{\phi}(\mathbf{x},t)$:

$$\boxed{\pi(\mathbf{x},t) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}(-i)\sqrt{\frac{\omega_{\mathbf{p}}}{2}}(A_{\mathbf{p}}e^{i(\mathbf{p}\cdot\mathbf{x}-\omega_{\mathbf{p}})t} - A^*_{\mathbf{p}}e^{-i(\mathbf{p}\cdot\mathbf{x}-\omega_{\mathbf{p}}t)})} \tag{12.1.10}$$

When we quantize these fields we will work in the Schrodinger picture, so the fields themselves will not be time-dependent. Consequently we can drop the time label and work solely in 3+0 space. It is now clear that:

$$\begin{cases}\tilde{\phi}(\mathbf{p})e^{i\mathbf{p}\cdot\mathbf{x}} = \tilde{\phi}(-\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{x}} = \frac{1}{\sqrt{2\omega_{\mathbf{p}}}}(A_{\mathbf{p}}e^{i\mathbf{p}\cdot\mathbf{x}} + A^*_{\mathbf{p}}e^{-i\mathbf{p}\cdot\mathbf{x}})\\ \tilde{\pi}(\mathbf{p})e^{i\mathbf{p}\cdot\mathbf{x}} = \tilde{\pi}(-\mathbf{p})e^{-i\mathbf{p}\cdot\mathbf{x}} = -i\sqrt{\frac{\omega_{\mathbf{p}}}{2}}(A_{\mathbf{p}}e^{i\mathbf{p}\cdot\mathbf{x}} - A^*_{\mathbf{p}}e^{-i\mathbf{p}\cdot\mathbf{x}})\end{cases} \tag{12.1.11}$$

$$\iff \begin{cases}A_{\mathbf{p}} = \sqrt{\frac{\omega_{\mathbf{p}}}{2}}(\tilde{\phi}(\mathbf{p}) + \frac{i}{\omega_{\mathbf{p}}}\tilde{\pi}(\mathbf{p}))\\ A^*_{\mathbf{p}} = \sqrt{\frac{\omega_{\mathbf{p}}}{2}}(\tilde{\phi}(-\mathbf{p}) - \frac{i}{\omega_{\mathbf{p}}}\tilde{\pi}(-\mathbf{p}))\end{cases} \tag{12.1.12}$$

Using (12.1.11) we can write the Klein-Gordon field Hamiltonian as:

$$\begin{aligned} H &= \frac{1}{2}\int d^3\mathbf{x}(\pi^2 + (\nabla\phi)^2 + m^2\phi^2) && (12.1.13)\\ &= \frac{1}{2}\int d^3\mathbf{x}\frac{d^3\mathbf{p}}{(2\pi)^3}\frac{d^3\mathbf{q}}{(2\pi)^3}\Big[-\frac{\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}{2}(A_{\mathbf{p}}e^{i\mathbf{p}\cdot\mathbf{x}} - A^*_{\mathbf{p}}e^{-i\mathbf{p}\cdot\mathbf{x}})(A_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{x}} - A^*_{\mathbf{q}}e^{-i\mathbf{q}\cdot\mathbf{x}}) && (12.1.14)\\ &\quad + \frac{1}{2\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}(i\mathbf{p}A_{\mathbf{p}}e^{i\mathbf{p}\cdot\mathbf{x}} - i\mathbf{p}A^*_{\mathbf{p}}e^{-i\mathbf{p}\cdot\mathbf{x}})\cdot(i\mathbf{q}A_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{x}} - i\mathbf{q}A^*_{\mathbf{q}}e^{-i\mathbf{q}\cdot\mathbf{x}}) && (12.1.15)\\ &\quad + \frac{m^2}{2\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}(A_{\mathbf{p}}e^{i\mathbf{p}\cdot\mathbf{x}} + A^*_{\mathbf{p}}e^{-i\mathbf{p}\cdot\mathbf{x}})(A_{\mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{x}} + A^*_{\mathbf{q}}e^{-i\mathbf{q}\cdot\mathbf{x}})\Big] && (12.1.16)\end{aligned}$$

This monstrosity simplifies a great deal, all thanks to Dirac and his delta function. Indeed note that when integrating over $\mathbf{x}$, the only relevant terms will be the exponentials. These will yield delta functions of the type:

$$\frac{1}{(2\pi)^3}e^{i\mathbf{p}\cdot\mathbf{x}}e^{i\mathbf{q}\cdot\mathbf{x}} \mapsto \delta^3(\mathbf{p}+\mathbf{q}),\ \frac{1}{(2\pi)^3}e^{-i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{x}} \mapsto \delta^3(\mathbf{p}+\mathbf{q}) \tag{12.1.17}$$

$$\frac{1}{(2\pi)^3}e^{i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{x}} \mapsto \delta^3(\mathbf{p}-\mathbf{q}),\ \frac{1}{(2\pi)^3}e^{-i\mathbf{p}\cdot\mathbf{x}}e^{i\mathbf{q}\cdot\mathbf{x}} \mapsto \delta^3(\mathbf{p}-\mathbf{q}) \tag{12.1.18}$$

Consequently:

$$H = \frac{1}{4}\int \frac{d^3\mathbf{p}d^3\mathbf{q}}{(2\pi)^3}\frac{1}{\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}\Big[(-\omega_{\mathbf{p}}\omega_{\mathbf{q}} - \mathbf{p}\cdot\mathbf{q} - m^2)(-A_{\mathbf{p}}A^*_{\mathbf{q}} - A^*_{\mathbf{p}}A_{\mathbf{q}})\delta^3(\mathbf{p}-\mathbf{q}) \tag{12.1.19}$$

$$+ (-\omega_{\mathbf{p}}\omega_{\mathbf{q}} - \mathbf{p}\cdot\mathbf{q} + m^2)(A_{\mathbf{p}}A_{\mathbf{q}} + A^*_{\mathbf{p}}A^*_{\mathbf{q}})\delta^3(\mathbf{p}+\mathbf{q})\Big] \tag{12.1.20}$$

$$= \frac{1}{4}\int \frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1}{\omega_{\mathbf{p}}}[(\omega^2_{\mathbf{p}} + \mathbf{p}^2 + m^2)(A_{\mathbf{p}}A^*_{\mathbf{p}} + A^*_{\mathbf{p}}A_{\mathbf{p}}) \tag{12.1.21}$$

$$+ (-\omega^2_{\mathbf{p}} + \mathbf{p}^2 + m^2)(A_{\mathbf{p}}A_{-\mathbf{p}} + A^*_{\mathbf{p}}A^*_{-\mathbf{p}})] \tag{12.1.22}$$

Recall however that $\omega^2_{\mathbf{p}} = \mathbf{p}^2 + m^2$ so the second term vanishes, giving:

$$H = \frac{1}{2}\int \frac{d^3\mathbf{p}}{(2\pi)^3}\,\omega_{\mathbf{p}}(A_{\mathbf{p}}A^*_{\mathbf{p}} + A^*_{\mathbf{p}}A_{\mathbf{p}}) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}\omega_{\mathbf{p}}|A_{\mathbf{p}}|^2 \tag{12.1.23}$$

Using (12.1.12) an immediate calculation finally yields:

$$\boxed{H = \frac{1}{2}\int \frac{d^3\mathbf{p}}{(2\pi)^3}[\omega^2_{\mathbf{p}}\tilde{\phi}_{\mathbf{p}}\tilde{\phi}_{-\mathbf{p}} + \tilde{\pi}_{\mathbf{p}}\tilde{\pi}_{-\mathbf{p}}]} \tag{12.1.24}$$

As expected, we get a bunch of independent harmonic oscillators!

The Quantum Oscillator

To quantize this classical field it will be useful to revisit some fundamental results about the quantum harmonic oscillator. The Hamiltonian operator reads:

$$H = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2 \tag{12.1.25}$$

We define the ladder operators:

$$a = \sqrt{\frac{\omega}{2}}q + \frac{i}{\sqrt{2\omega}}p,\ a^\dagger = \sqrt{\frac{\omega}{2}}q - \frac{i}{\sqrt{2\omega}}p \tag{12.1.26}$$

or alternatively:

$$q = \frac{1}{\sqrt{2\omega}}(a + a^\dagger),\ p = -i\sqrt{\frac{\omega}{2}}(a - a^\dagger) \tag{12.1.27}$$

Using the canonical commutation rule we find:

$$[p, q] = i \implies [a, a^\dagger] = 1 \tag{12.1.28}$$

and the Hamiltonian now reads:

$$H = \omega\left(a^\dagger a + \frac{1}{2}\right) \tag{12.1.29}$$

It can easily be shown that:

$$[H, a^\dagger] = \omega a^\dagger,\ [H, a] = -\omega a \tag{12.1.30}$$

implying that given an eigenstate $|E\rangle$ with energy E then:

$$H a^\dagger |E\rangle = (E + \omega) a^\dagger |E\rangle\,,\ H a |E\rangle = (E - \omega) a |E\rangle \tag{12.1.31}$$

The spectrum of the Hamiltonian thus consists of a ladder of energy levels with spacing ω. We must have a lower bound to the spectrum, so given a ground state $|0\rangle$ then we must require $a|0\rangle = 0$ and thus:

$$H|0\rangle = \frac{1}{2}\omega|0\rangle \tag{12.1.32}$$

Finally, defining $|n\rangle = (a^\dagger)^n |0\rangle$ then

$$\hat{H}|n\rangle = \left(n + \frac{1}{2}\right)\omega|n\rangle \tag{12.1.33}$$

Quantizing the Klein-Gordon field

Returning to the Klein-Gordon equation, we can promote $\phi(\mathbf{x})$ and $\phi(\mathbf{x})$ to operator-valued fields, **quantum fields**. As a result $A_\mathbf{p}$ and $A^*_\mathbf{p}$ will be promoted to operators $a_\mathbf{p}$ and $a^\dagger_\mathbf{p}$, defined as:

$$a_\mathbf{p} = \sqrt{\frac{\omega_\mathbf{p}}{2}}\left(\tilde{\phi}(\mathbf{p}) + \frac{i}{\omega_\mathbf{p}}\tilde{\pi}(\mathbf{p})\right) \tag{12.1.34a}$$

$$a^\dagger_\mathbf{p} = \sqrt{\frac{\omega_\mathbf{p}}{2}}\left(\tilde{\phi}(\mathbf{p}) - \frac{i}{\omega_\mathbf{p}}\tilde{\pi}(\mathbf{p})\right) \tag{12.1.34b}$$

completely analogously to $A_\mathbf{p}$ and $A^*_\mathbf{p}$ in (12.1.12). This yields the following expression for the quantum fields (note importantly that these fields are operator valued):

$$\phi(\mathbf{x}) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1}{\sqrt{2\omega_\mathbf{p}}}(a_\mathbf{p} e^{i\mathbf{p}\cdot\mathbf{x}} + a^\dagger_\mathbf{p} e^{-i\mathbf{p}\cdot\mathbf{x}}) \tag{12.1.35a}$$

$$\pi(\mathbf{x}) = \int \frac{d^3\mathbf{p}}{(2\pi)^3}(-i)\sqrt{\frac{\omega_\mathbf{p}}{2}}(a_\mathbf{p} e^{i\mathbf{p}\cdot\mathbf{x}} - a^\dagger_\mathbf{p} e^{-i\mathbf{p}\cdot\mathbf{x}}) \tag{12.1.35b}$$

We see that the canonical commutation rules for the fields are equivalent to the canonical commutation rules for the ladder operators:

$$\begin{cases}[\phi(\mathbf{x}),\phi(\mathbf{y})]=[\pi(\mathbf{x}),\pi(\mathbf{y})]=0\\ [\phi(\mathbf{x}),\pi(\mathbf{y})]=i\delta^3(\mathbf{x}-\mathbf{y})\end{cases}\iff\begin{cases}[a_{\mathbf{p}},a_{\mathbf{q}}]=[a^\dagger_{\mathbf{p}},a^\dagger_{\mathbf{q}}]=0\\ [a_{\mathbf{p}},a^\dagger_{\mathbf{q}}]=(2\pi)^3\delta^3(\mathbf{p}-\mathbf{q})\end{cases}\tag{12.1.36}$$

Proof. We prove this in the $\Longrightarrow$ direction (the other way is more of the same stuff). It is easiest to first derive the commutation rules for $\tilde{\phi}(\mathbf{p})$ and $\tilde{\pi}(\mathbf{p})$. We have that:

$$[\tilde{\phi}(\mathbf{p}),\tilde{\pi}(\mathbf{q})]=\int d^3\mathbf{x}d^3\mathbf{y}e^{-i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{y}}[\phi(\mathbf{x}),\pi(\mathbf{y})]\tag{12.1.37}$$

$$=\int d^3\mathbf{x}d^3\mathbf{y}e^{-i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{y}}i\delta^3(\mathbf{x}-\mathbf{y})=i(2\pi)^3\delta(\mathbf{p}+\mathbf{q})\tag{12.1.38}$$

where we used the linearity of $[\cdot,\cdot]$. This is an interesting result, it tells us that the conjugate operator to $\tilde{\phi}(\mathbf{p})$ is not $\tilde{\pi}(\mathbf{p})$ but rather $\tilde{\pi}(-\mathbf{p})$. Similarly:

$$[\tilde{\phi}(\mathbf{p}),\tilde{\phi}(\mathbf{q})]=\int d^3\mathbf{x}d^3\mathbf{y}e^{-i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{y}}[\phi(\mathbf{x}),\phi(\mathbf{y})]=0\tag{12.1.39}$$

$$[\tilde{\pi}(\mathbf{p}),\tilde{\pi}(\mathbf{q})]=\int d^3\mathbf{x}d^3\mathbf{y}e^{-i\mathbf{p}\cdot\mathbf{x}}e^{-i\mathbf{q}\cdot\mathbf{y}}[\pi(\mathbf{x}),\pi(\mathbf{y})]=0\tag{12.1.40}$$

Therefore, we find that:

$$[a_{\mathbf{p}},a_{\mathbf{q}}]=\frac{\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}{2}\left(\frac{i}{\omega_{\mathbf{p}}}[\tilde{\pi}_{\mathbf{p}},\tilde{\phi}_{-\mathbf{q}}]-\frac{i}{\omega_{\mathbf{q}}}[\tilde{\phi}_{\mathbf{p}},\tilde{\pi}_{-\mathbf{q}}]\right)\tag{12.1.41}$$

$$=\frac{i\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}{2}\cdot(2\pi)^3 i\left(-\frac{\delta^3(\mathbf{p}-\mathbf{q})}{\omega_{\mathbf{p}}}-\frac{\delta^3(\mathbf{p}-\mathbf{q})}{\omega_{\mathbf{p}}}\right)\tag{12.1.42}$$

$$=(2\pi)^3\delta^3(\mathbf{p}-\mathbf{q})\tag{12.1.43}$$

as we wished to prove. Similarly we find:

$$[a_{\mathbf{p}},a_{\mathbf{q}}]=\frac{\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}{2}\left(\frac{i}{\omega_{\mathbf{q}}}[\tilde{\phi}_{\mathbf{p}},\tilde{\pi}_{\mathbf{q}}]+\frac{i}{\omega_{\mathbf{p}}}[\tilde{\pi}_{\mathbf{p}},\tilde{\phi}_{\mathbf{q}}]\right)=0\tag{12.1.44}$$

$$[a^\dagger_{\mathbf{p}},a^\dagger_{\mathbf{q}}]=-\frac{\sqrt{\omega_{\mathbf{p}}\omega_{\mathbf{q}}}}{2}\left(\frac{i}{\omega_{\mathbf{q}}}[\tilde{\phi}_{-\mathbf{p}},\tilde{\pi}_{-\mathbf{q}}]+\frac{i}{\omega_{\mathbf{p}}}[\tilde{\pi}_{-\mathbf{p}},\tilde{\phi}_{-\mathbf{q}}]\right)=0 \qquad \blacksquare\tag{12.1.45}$$

Now the Hamiltonian in (12.1.23) becomes a Hamiltonian operator expressed as:

$$H=\frac{1}{2}\int\frac{d^3\mathbf{p}}{(2\pi)^3}\,\omega_{\mathbf{p}}(a_{\mathbf{p}}a^\dagger_{\mathbf{p}}+a^\dagger_{\mathbf{p}}a_{\mathbf{p}})\tag{12.1.46}$$

Using the commutation rules in (12.1.36) this may be written in a more suitable form:

$$\boxed{H=\int\frac{d^3\mathbf{p}}{(2\pi)^3}\,\omega_{\mathbf{p}}\left(a^\dagger_{\mathbf{p}}a_{\mathbf{p}}+\frac{1}{2}(2\pi)^3\delta^3(0)\right)}\tag{12.1.47}$$

Note that a each momentum mode evolves independently, there are no interactions between different $\mathbf{p}$'s so we do indeed have a free field theory. One worrying term however is the delta function which we are evaluating at zero, the only point where it is defined to be infinitely large, and we do not want infinities in our theory. Even worse, we are integrating this infinity over all our degrees of freedom, which are uncountably infinite.

12.2 Infinities in the vacuum

The vacuum state

Define the vacuum state $|0\rangle$ to be such that:

$$a_{\mathbf{p}} |0\rangle = 0, \ \forall \mathbf{p} \tag{12.2.1}$$

Applying our Hamiltonian on this state we find:

$$H |0\rangle = \int \frac{d^3\mathbf{p}}{(2\pi)^3} \frac{1}{2} (2\pi)^3 \delta^3(0) |0\rangle \stackrel{?!}{=} \infty |0\rangle \tag{12.2.2}$$

As we said earlier, there are two infinities in this result: one coming from the infinite number of degrees of freedom (infra-red divergences due to the large length scale), and one from the delta function.

Thus, let us consider a box of size L. Trivially

$$(2\pi)^3 \delta^3(0) = \lim_{L\to\infty} \int_{-L/2}^{L/2} d^3\mathbf{x}\, e^{i\mathbf{p}\cdot\mathbf{x}} \Big|_{\mathbf{p}=0} = L^3 \tag{12.2.3}$$

so in a finite box the delta function could have been replaced by the volume of the box. Consequently the ground state energy density is:

$$\varepsilon_0 = \int \frac{d^3\mathbf{p}}{(2\pi)^3} \frac{\omega_{\mathbf{p}}}{2} \tag{12.2.4}$$

This integral is still infinite as $\mathbf{p} \to \infty$, that is at infinitely small wavelengths (UV divergence). However, we should not expect our solution to hold for arbitrarily small length scales [1], so we should impose an energy cut-off to our integral. For example, in condensed matter theory we often deal with discrete lattices, so the minimal length scale to be considered is the lattice spacing.

More practically, since in experiments we can only really measure energy differences, we can ignore delta function and simply write:

$$\boxed{H = \int \frac{d^3\mathbf{p}}{(2\pi)^3} \omega_{\mathbf{p}} a^\dagger_{\mathbf{p}} a_{\mathbf{p}}} \tag{12.2.5}$$

This is equivalent to redefining our hamiltonian so as to remove the delta function. Note that now the zero point-energy is equal to 0. For example, we could have written the

[1] just like we would not expect classical electromagnetism to hold at quantum scales where Coulomb's law diverges

classical hamiltonian as $H = \frac{1}{2}(\omega q - ip)(\omega q + ip)$. This however would give a quantum hamiltonian $\hat{H} = \omega a^\dagger a$, so there is an ambiguity in the quantization process due to the fact that while classical observables commute, quantum operators do not. To deal with this we can set a convention, namely **normal ordering** which places annihilation operators to the right of creation operators, and (12.2.5) would be written as:

$$: H := \int \frac{d^3\mathbf{p}}{(2\pi)^3} \omega_{\mathbf{p}} a^\dagger_{\mathbf{p}} a_{\mathbf{p}} \tag{12.2.6}$$

It is now easy to check that:

$$[H, a^\dagger_{\mathbf{p}}] = \omega_{\mathbf{p}} a^\dagger_{\mathbf{p}},\ [H, a_{\mathbf{p}}] = -\omega_{\mathbf{p}} a_{\mathbf{p}} \tag{12.2.7}$$

The Casimir effect

12.3 Particles from fields

Let us define $|\mathbf{p}\rangle = a^\dagger_{\mathbf{p}} |0\rangle$, so that $H|p\rangle = \omega_{\mathbf{p}} |\mathbf{p}\rangle = E_{\mathbf{p}} |\mathbf{p}\rangle$. It follows that:

$$E^2_{\mathbf{p}} = \mathbf{p}^2 + m^2 \tag{12.3.1}$$

which is the relativistic dispersion relation for a massive particle with momentum $\mathbf{p}$. Thus we should interpret $|\mathbf{p}\rangle$ as the state of one such particle. So the coefficient of ϕ^2 in the KG field became a mass, and the frequencies decomposition became momenta!

Since $|\mathbf{p}\rangle$ is a momentum state (plane wave), we would like to have a momentum operator to give us $\mathbf{p}$ when acting on this state. In classical field theory we defined the momentum of a field as:

$$p^i = \int d^3\mathbf{x} T^{0i} \tag{12.3.2}$$

which for the Klein-Gordon field reads:

$$\mathbf{p} = -\int d^3\mathbf{x} \pi(\mathbf{x}) \nabla\phi(\mathbf{x}) \tag{12.3.3}$$

which upon quantization turns into the operator:

$$\mathbf{p} = \int \frac{d^3\mathbf{p}}{(2\pi)^3} \mathbf{p} a^\dagger_{\mathbf{p}} a_{\mathbf{p}} \tag{12.3.4}$$

Note that:

$$\mathbf{p}|\mathbf{q}\rangle = \int \frac{d^3\mathbf{p}}{(2\pi)^3} \mathbf{p} a^\dagger_{\mathbf{p}} a_{\mathbf{p}} |\mathbf{q}\rangle = \int \frac{d^3\mathbf{p}}{(2\pi)^3} \mathbf{p}\delta(\mathbf{p}-\mathbf{q})|\mathbf{q}\rangle = \mathbf{q}|\mathbf{q}\rangle \tag{12.3.5}$$

as desired.

Similarly, we can also define an angular momentum operator:

$$J^i = \epsilon^{ijk} \int d^3\mathbf{x} (M^0)^j k \tag{12.3.6}$$

12.4 Quantizing the electromagnetic field

A very similar quantization process can be performed for vector fields, most importantly, the electromagnetic field.

For non-relativistic systems we typically use the Coulomb gauge $\nabla \cdot \mathbf{A} = 0$, so that in vacuum the electric and magnetic fields are given by:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{12.4.1}$$

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} \tag{12.4.2}$$

Inserting these into the Ampere-Maxwell law we find that:

$$\nabla^2 \mathbf{A} = \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} \tag{12.4.3}$$

which is the classical wave-equation. If we take $\mathbf{A}$ to be in a box of volume $\mathcal{V}$ with periodic boundary conditions, then the solutions to the above will be those of a waveguide, and can thus be expanded into modes:

$$\mathbf{A}(\mathbf{x}, t) = \sum_{\mathbf{k}} \mathbf{A}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{12.4.4}$$

which when substituted into (12.4.3) yields:

$$\mathbf{A}(\mathbf{x}, t) = \sum_{\mathbf{k}} (\mathbf{A}^+_{\mathbf{k}} e^{i\omega_{\mathbf{k}} t} + \mathbf{A}^-_{\mathbf{k}} e^{-i\omega_{\mathbf{k}} t}) e^{i\mathbf{k}\cdot\mathbf{x}} \tag{12.4.5}$$

$$= \sum_{\mathbf{k}} (\mathbf{A}^+_{-\mathbf{k}} e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)} + \mathbf{A}^-_{\mathbf{k}} e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)}) \tag{12.4.6}$$

Since the vector potential must be a real quantity, we must have that:

$$\sum_{\mathbf{k}} (\mathbf{A}^+_{-\mathbf{k}} e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)} + \mathbf{A}^-_{\mathbf{k}} e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)}) = \sum_{\mathbf{k}} ((\mathbf{A}^+_{-\mathbf{k}})^* e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)} + (\mathbf{A}^-_{\mathbf{k}})^* e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)}) \tag{12.4.7}$$

so that $\mathbf{A}^+_{-\mathbf{k}} = (\mathbf{A}^-_{\mathbf{k}})^*$. It follows that we may decompose the vector potential into modes of wave-vector $\mathbf{k}$ and polarisation $\boldsymbol{\epsilon}_\lambda$ by letting $\mathbf{A}^-_{\mathbf{k}} = A_{\mathbf{k},\lambda}\boldsymbol{\epsilon}_\lambda$:

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\mathbf{k}} \sum_{\lambda=1,2} (A_{\mathbf{k},\lambda} e^{i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)} + A^*_{\mathbf{k},\lambda} e^{-i(\mathbf{k}\cdot\mathbf{x} - \omega_{\mathbf{k}} t)}) \boldsymbol{\epsilon}_{\mathbf{k},\lambda} \tag{12.4.8}$$

where $\{\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \mathbf{k}/|\mathbf{k}|\}$ form an orthonormal basis and $\omega_{\mathbf{k}} = |\mathbf{k}|c$. The classical hamiltonian for the electromagnetic field is given by

$$\hat{H} = \frac{1}{2}\int (\varepsilon_0 |\mathbf{E}|^2 + \frac{1}{\mu_0}|\mathbf{B}|^2) d\mathbf{x} = \frac{1}{2}\int \left(\varepsilon_0 \left|\frac{\partial \mathbf{A}}{\partial t}\right|^2 + \frac{1}{\mu_0}|\nabla \times \mathbf{A}|^2\right) d\mathbf{x} \tag{12.4.9}$$

We now use some Fourier analysis trickery to simplify the above expression. Firstly, note that by applying Parseval's theorem (not to be confused with Parseval's *identity*), which states that:

$$\int |f(\mathbf{x})|^2 d\mathbf{x} = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\mathbf{k}} |\widetilde{f}(\mathbf{k})|^2 \tag{12.4.10}$$

then we get (we ignore any prefactors in front of the sum as we will normalize everything at the end):

$$\int \left|\frac{\partial \mathbf{A}}{\partial t}\right|^2 d\mathbf{x} = \sum_{\mathbf{k}} \left|\mathcal{F}\left(\frac{\partial \mathbf{A}}{\partial t}\right)\right|^2, \int |\nabla\times\mathbf{A}|^2 d\mathbf{x} = \sum_{\mathbf{k}} \left|\mathcal{F}(\nabla\times\mathbf{A})\right|^2 \tag{12.4.11}$$

Computing the Fourier transform is immediate:

$$\mathcal{F}(\nabla\times\mathbf{A})(\mathbf{k}') = \sum_{\mathbf{k},\lambda} (i\mathbf{k}\times\boldsymbol{\epsilon}_{\mathbf{k},\lambda})(A_{\mathbf{k},\lambda}e^{-i\omega_{\mathbf{k}}t}\delta_{\mathbf{k}'} - A^*_{\mathbf{k},\lambda}e^{i\omega_{\mathbf{k}}t}\delta_{-\mathbf{k}'}) \tag{12.4.12}$$

$$\mathcal{F}\left(\frac{\partial \mathbf{A}}{\partial t}\right)(\mathbf{k}') = \sum_{\mathbf{k},\lambda} (i\omega_{\mathbf{k}})(A_{\mathbf{k},\lambda}e^{-i\omega_{\mathbf{k}}t}\delta_{\mathbf{k}'} - A^*_{\mathbf{k},\lambda}e^{i\omega_{\mathbf{k}}t}\delta_{-\mathbf{k}'})\boldsymbol{\epsilon}_{\mathbf{k},\lambda} \tag{12.4.13}$$

where $\delta_{\mathbf{k}'}$ is shorthand for $\delta_{\mathbf{k}',\mathbf{k}}$. We also note that:

$$\boldsymbol{\epsilon}_{\mathbf{k},\lambda}\cdot\boldsymbol{\epsilon}_{t\mathbf{k},\epsilon'_\lambda} = \delta_{\lambda,\lambda'} \tag{12.4.14}$$

and[2]

$$(i\mathbf{k}\times\boldsymbol{\epsilon}_\lambda)\cdot(-i\mathbf{k}\times\boldsymbol{\epsilon}_{\lambda'}) = |\mathbf{k}|^2\delta_{\lambda,\lambda'} \tag{12.4.15}$$

so that

$$\sum_{\mathbf{k}'} |\mathcal{F}(\nabla\times\mathbf{A})|^2 = \sum_{\substack{\mathbf{k},\mathbf{k}'\\ \lambda,\lambda'}} (i\mathbf{k}\times\boldsymbol{\epsilon}_\lambda)\cdot(-i\mathbf{k}\times\boldsymbol{\epsilon}_{\lambda'})(A_{\mathbf{k},\lambda}e^{-i\omega_{\mathbf{k}}t}\delta_{\mathbf{k}'} - A^*_{\mathbf{k},\lambda}e^{i\omega_{\mathbf{k}}t}\delta_{-\mathbf{k}'}) \tag{12.4.16}$$

$$\times\,(A^*_{\mathbf{k},\lambda'}e^{i\omega_{\mathbf{k}}t}\delta_{\mathbf{k}'} - A_{\mathbf{k},\lambda'}e^{-i\omega_{\mathbf{k}}t}\delta_{-\mathbf{k}'}) \tag{12.4.17}$$

$$= \sum_{\mathbf{k},\lambda} |\mathbf{k}|^2(2|A_{\mathbf{k},\lambda}|^2 - A_{\mathbf{k},\lambda}A_{\mathbf{k},\lambda}e^{-2i\omega_{\mathbf{k}}t} - A^*_{\mathbf{k},\lambda}A^*_{\mathbf{k},\lambda}e^{2i\omega_{\mathbf{k}}t}) \tag{12.4.18}$$

and similarly:

$$\sum_{\mathbf{k}'} \left|\mathcal{F}\left(\frac{\partial \mathbf{A}}{\partial t}\right)\right|^2 = \sum_{\substack{\mathbf{k},\mathbf{k}'\\ \lambda,\lambda'}} (i\omega)(-i\omega)(A_{\mathbf{k},\lambda}e^{-i\omega_{\mathbf{k}}t} - A^*_{\mathbf{k},\lambda}e^{i\omega_{\mathbf{k}}t})(A_{\mathbf{k},\lambda}e^{-i\omega_{\mathbf{k}}t} - A^*_{\mathbf{k},\lambda}e^{i\omega_{\mathbf{k}}t})\boldsymbol{\epsilon}_\lambda\cdot\boldsymbol{\epsilon}_{\lambda'}$$

$$= \sum_{\mathbf{k},\lambda} \omega_{\mathbf{k}}^2(2|A_{\mathbf{k},\lambda}|^2 + A_{\mathbf{k},\lambda}A_{\mathbf{k},\lambda}e^{-2i\omega_{\mathbf{k}}t} + A^*_{\mathbf{k},\lambda}A^*_{\mathbf{k},\lambda}e^{2i\omega_{\mathbf{k}}t})$$

[2] this is easy to prove:

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{a}\times\mathbf{c}) &= \epsilon_{ijk}a_jb_k\epsilon_{imn}a_mc_n\\ &= (\delta_{jm}\delta_{kn} - \delta_{jn}\delta_{km})a_ja_mb_kc_n\\ &= |\mathbf{a}|^2(\mathbf{b}\cdot\mathbf{c}) - (\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{b})\end{aligned}$$

and since $\mathbf{b} = \epsilon_\lambda$ is orthonormal to $\mathbf{c} = \epsilon_{\lambda'}$, and since $\mathbf{k}$ is orthogonal to both polarization vectors, we get the desired, result.

Finally, we find that (we switch $\mathbf{k}' \to \mathbf{k}$ for convenience):

$$\hat{H} = 2\varepsilon_0 \sum_{\mathbf{k},\lambda} \omega_{\mathbf{k}}^2 |A_{\mathbf{k},\lambda}|^2 = 2\varepsilon_0 \sum_{\mathbf{k},\lambda} \omega_{\mathbf{k}}^2 (|A_{\mathbf{k},\lambda}^R|^2 + |A_{\mathbf{k},\lambda}^I|^2) \tag{12.4.19}$$

where $A_{\mathbf{k},\lambda} = A_{\mathbf{k},\lambda}^R + iA_{\mathbf{k},\lambda}^I$. Note that if we define $A_{\mathbf{k},\lambda}(t) = A_{\mathbf{k},\lambda} e^{-i\omega_{\mathbf{k}} t}$ then:

$$\dot{A}_{\mathbf{k},\lambda}^R = \omega_{\mathbf{k}} A_{\mathbf{k},\lambda}^I, \ \dot{A}_{\mathbf{k},\lambda}^I = -\omega_{\mathbf{k}} A_{\mathbf{k},\lambda}^R \tag{12.4.20}$$

and thus:

$$\frac{\partial H}{\partial A_{\mathbf{k},\lambda}^R} = 4\varepsilon_0 \omega_{\mathbf{k}}^2 A_{\mathbf{k},\lambda}^R = -4\varepsilon_0 \omega_{\mathbf{k}} \dot{A}_{\mathbf{k},\lambda}^I \tag{12.4.21}$$

$$\frac{\partial H}{\partial A_{\mathbf{k},\lambda}^I} = 4\varepsilon_0 \omega_{\mathbf{k}}^2 A_{\mathbf{k},\lambda}^I = 4\varepsilon_0 \omega_{\mathbf{k}} \dot{A}_{\mathbf{k},\lambda}^R \tag{12.4.22}$$

implying that $A_{\mathbf{k},\lambda}^R$ and $A_{\mathbf{k},\lambda}^I$ are canonically conjugate variables (up to some proportionality constant). So, we may define the conjugate position and conjugate momenta to be:

$$Q_{\mathbf{k},\lambda} = 2\sqrt{\varepsilon_0} A_{\mathbf{k},\lambda}^R \tag{12.4.23}$$

$$P_{\mathbf{k},\lambda} = 2\omega_{\mathbf{k}} \sqrt{\varepsilon_0} A_{\mathbf{k},\lambda}^I \tag{12.4.24}$$

respectively. Clearly, these satisfy:

$$\begin{cases} \dot{Q}_{\mathbf{k},\lambda} = P_{\mathbf{k},\lambda} \\ \dot{P}_{\mathbf{k},\lambda} = -\omega_{\mathbf{k}}^2 P_{\mathbf{k},\lambda} \end{cases} \qquad \begin{cases} \frac{\partial H}{\partial Q_{\mathbf{k},\lambda}} = -\dot{P}_{\mathbf{k},\lambda} \\ \frac{\partial H}{\partial P_{\mathbf{k},\lambda}} = \dot{Q}_{\mathbf{k},\lambda} \end{cases} \tag{12.4.25}$$

as would be the case for a harmonic oscillator. Consequently, also the hamiltonian will be identical to that of a harmonic oscillator:

$$H = \frac{1}{2} \sum_{\mathbf{k},\lambda} (P_{\mathbf{k},\lambda}^2 + \omega_{\mathbf{k}}^2 Q_{\mathbf{k},\lambda}) \tag{12.4.26}$$

We can now quantize the electromagnetic field just as one would quantize the harmonic oscillator. We promote $P_{\mathbf{k},\lambda}$ and $Q_{\mathbf{k},\lambda}$ to quantum operators $\hat{p}_{\mathbf{k},\lambda}$ and $\hat{q}_{\mathbf{k},\lambda}$ which satisfy the canonical commutation relations:

$$[\hat{q}_{\mathbf{k},\lambda}, \hat{p}_{\mathbf{k}',\lambda'}] = i\hbar \delta_{\mathbf{k}\mathbf{k}'} \delta_{\lambda\lambda'} \tag{12.4.27}$$

$$[\hat{q}_{\mathbf{k},\lambda}, \hat{q}_{\mathbf{k}',\lambda'}] = [\hat{p}_{\mathbf{k},\lambda}, \hat{p}_{\mathbf{k}',\lambda'}] = 0 \tag{12.4.28}$$

so that:

$$\hat{H} = \frac{1}{2} \sum_{\mathbf{k},\lambda} (\hat{p}_{\mathbf{k},\lambda}^2 + \omega_{\mathbf{k}}^2 \hat{q}_{\mathbf{k},\lambda}^2) \tag{12.4.29}$$

We introduce the ladder operators:

$$\begin{cases} a_{\mathbf{k},\lambda}^\dagger = \sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}} (\omega_{\mathbf{k}} \hat{q}_{\mathbf{k},\lambda} - i\hat{p}_{\mathbf{k},\lambda}) \\ a_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}} (\omega_{\mathbf{k}} \hat{q}_{\mathbf{k},\lambda} + i\hat{p}_{\mathbf{k},\lambda}) \end{cases} \Longrightarrow \begin{cases} \hat{q}_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}} (a_{\mathbf{k},\lambda}^\dagger + a_{\mathbf{k},\lambda}) \\ \hat{p}_{\mathbf{k},\lambda} = i\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2}} (a_{\mathbf{k},\lambda}^\dagger - a_{\mathbf{k},\lambda}) \end{cases} \tag{12.4.30}$$

With these new operators, the Hamiltonian turns into the familiar quantum harmonic oscillator:

$$\hat{H} = \hbar\omega_{\mathbf{k}} \sum_{\mathbf{k},\lambda} \left(a^\dagger_{\mathbf{k},\lambda} a_{\mathbf{k},\lambda} + \frac{1}{2} \right) \tag{12.4.31}$$

Finally, to relate this hamiltonian to our classical expression (12.4.8) of the vector potential, we make use of the fact that:

$$A_{\mathbf{k},\lambda} = \frac{1}{2\sqrt{\varepsilon_0}} \sqrt{\frac{\hbar}{2\omega_{\mathbf{k}}}} (a^\dagger_{\mathbf{k},\lambda} + a_{\mathbf{k},\lambda}) + \frac{i}{2\sqrt{\varepsilon_0}\omega_{\mathbf{k}}} i \sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2}} (a^\dagger_{\mathbf{k},\lambda} - a_{\mathbf{k},\lambda}) \tag{12.4.32}$$

$$= \sqrt{\frac{\hbar}{2\varepsilon_0\omega_{\mathbf{k}}}} a_{\mathbf{k},\lambda} \implies A^*_{\mathbf{k},\lambda} = \sqrt{\frac{\hbar}{2\varepsilon_0\omega_{\mathbf{k}}}} a^\dagger_{\mathbf{k},\lambda} \tag{12.4.33}$$

giving:

$$\mathbf{A}(\mathbf{x}, t) = \sqrt{\frac{\hbar}{2\varepsilon_0\omega_{\mathbf{k}}\mathcal{V}}} \sum_{\mathbf{k}} \sum_{\lambda=1,2} (e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} a_{\mathbf{k},\lambda} + e^{-i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} a^\dagger_{\mathbf{k},\lambda}) \boldsymbol{\epsilon}_\lambda \tag{12.4.34}$$

$$\mathbf{E}(\mathbf{x}, t) = i\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0\mathcal{V}}} \sum_{\mathbf{k}} \sum_{\lambda=1,2} (e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} a_{\mathbf{k},\lambda} - e^{-i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} a^\dagger_{\mathbf{k},\lambda}) \boldsymbol{\epsilon}_\lambda \tag{12.4.35}$$

$$\mathbf{B}(\mathbf{x}, t) = i\sqrt{\frac{\hbar}{2\varepsilon_0\omega_{\mathbf{k}}\mathcal{V}}} \sum_{\mathbf{k}} \sum_{\lambda=1,2} (e^{i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} a_{\mathbf{k},\lambda} - e^{-i(\mathbf{k}\cdot\mathbf{x}-\omega_{\mathbf{k}}t)} a^\dagger_{\mathbf{k},\lambda}) (\mathbf{k} \times \boldsymbol{\epsilon}_\lambda) \tag{12.4.36}$$

12.5 Quantizing a complex scalar field

We have discussed real scalar and vector fields, so it is now time to tackle complex scalar fields.

Second quantization

13

13.1 The need for second quantization

Suppose we have an N-particle system, where particle i resides in a hilbert space $\mathcal{H}_i$. The system as a whole will then be described by a state in the tensor product space $\bigotimes_{i=1}^{n} \mathcal{H}_i$. In the special case where the N-particles are indistinguishable, special care must be made due to the distinction between fermions and bosons. The states describing bosons will be totally symmetric under particle exchange, and thus belong to the subspace $\mathrm{Sym}^N \mathcal{H}$ while states describing bosons will be totally anti-symmetric, and belong to the subspace $\bigwedge^N \mathcal{H}$.

Let $|\psi\rangle = |\psi^{(1)}\rangle_1 \otimes |\psi^{(2)}\rangle_2 \otimes ... |\psi^{(N)}\rangle_N \in \mathcal{H}^N$. This doesn't automatically qualify $|\psi\rangle$ as a physical state describing bosonic or fermionic systems. We must find a way to symmetrize or anti-symmetrize this state. It can be shown that this can be done through the projection operators:

$$\hat{S}_+ = \frac{1}{\sqrt{N!}} \sum_{\sigma \in S_N} \hat{P}_\sigma \tag{13.1.1}$$

$$\hat{S}_- = \frac{1}{\sqrt{N!}} \sum_{\sigma \in S_N} \mathrm{sgn}(\sigma) \hat{P}_\sigma \tag{13.1.2}$$

known as the symmetrization and anti-symmetrization operators. Using the definition of permanents (denoted by a + sign at the top) and determinants, it follows that:

$$|\psi\rangle_+ = \hat{S}_+ |\psi\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} |\psi^{(1)}\rangle_1 & |\psi^{(1)}\rangle_2 & |\psi^{(1)}\rangle_3 & \cdots & |\psi^{(1)}\rangle_N \\ |\psi^{(2)}\rangle_1 & |\psi^{(2)}\rangle_2 & |\psi^{(2)}\rangle_3 & \cdots & |\psi^{(2)}\rangle_N \\ |\psi^{(3)}\rangle_1 & |\psi^{(3)}\rangle_2 & |\psi^{(3)}\rangle_3 & \cdots & |\psi^{(3)}\rangle_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ |\psi^{(N)}\rangle_1 & |\psi^{(N)}\rangle_2 & |\psi^{(N)}\rangle_3 & \cdots & |\psi^{(N)}\rangle_N \end{vmatrix}^+ \tag{13.1.3}$$

and similarly:

$$|\psi\rangle_- = \hat{S}_- |\psi\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} |\psi^{(1)}\rangle_1 & |\psi^{(1)}\rangle_2 & |\psi^{(1)}\rangle_3 & \cdots & |\psi^{(1)}\rangle_N \\ |\psi^{(2)}\rangle_1 & |\psi^{(2)}\rangle_2 & |\psi^{(2)}\rangle_3 & \cdots & |\psi^{(2)}\rangle_N \\ |\psi^{(3)}\rangle_1 & |\psi^{(3)}\rangle_2 & |\psi^{(3)}\rangle_3 & \cdots & |\psi^{(3)}\rangle_N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ |\psi^{(N)}\rangle_1 & |\psi^{(N)}\rangle_2 & |\psi^{(N)}\rangle_3 & \cdots & |\psi^{(N)}\rangle_N \end{vmatrix} \tag{13.1.4}$$

We can write these results more intuitively as:

$$|\psi\rangle_+ = \frac{1}{\sqrt{N!}}(|\psi\rangle + \text{permutations of } |\psi\rangle) \tag{13.1.5}$$

$$|\psi\rangle_- = \frac{1}{\sqrt{N!}}(|\psi\rangle \pm \text{permutations of } |\psi\rangle) \tag{13.1.6}$$

To summarize, we started with some ket where a list of N states in $\mathcal{H}$ were occupied by a particle, and produced a new state where each state is still occupied, but that is now (anti-)invariant under any particle exchange. We have gone from thinking about the state of each particle to thinking about which states are occupied.

It is clear that calculations involving permanents and determinants can get very messy in the thermodynamic limit, due to the $\sim o(N!)$ complexity of evaluating determinants and permanents. A new convention is thus needed to deal with many-body systems such as the ones encountered in condensed matter systems.

The situation is further worsened by a redundancy in the standard notation we have used thus far. Consider the following states:

$$|\Psi_1\rangle = |\psi^{(1)}\rangle_1 \otimes |\psi^{(2)}\rangle_2 \otimes |\psi^{(3)}\rangle_3 \otimes |\psi^{(4)}\rangle_4 \tag{13.1.7}$$

$$|\Psi_2\rangle = |\psi^{(4)}\rangle_1 \otimes |\psi^{(2)}\rangle_2 \otimes |\psi^{(1)}\rangle_3 \otimes |\psi^{(3)}\rangle_4 \tag{13.1.8}$$

$$\tag{13.1.9}$$

It is clear that (anti)-symmetrizing $|\Psi_1\rangle$ and $|\Psi_2\rangle$ will give the same state. More generally, for fermionic systems, given any state in $\mathcal{H}^N$, there will be $N!$ states generated by the symmetric group S_N which get symmetrized to the same state in $\bigwedge^N \mathcal{H}$. [1]In other words, the dimension of $\mathcal{H}^N$ does not match the dimensions of $\mathrm{Sym}_N\mathcal{H}$ and $\bigwedge^N \mathcal{H}$.

13.2 The occupation representation and Fock spaces

One important concept that came up in the previous section was the occupation of states. Indeed, in both the symmetrized and anti-symmetrized states, the occupation of each state was preserved. This suggests using a notation where instead of referring which particle occupies which state, we refer to which states are occupied. This is known as the **occupation representation**.

Generally, if we let $\{|\psi^{(1)}\rangle, |\psi^{(2)}\rangle, ..., |\psi^{(k)}, ...\rangle\}$ be an ordered basis of $\mathcal{H}$, then we define

$$|n_1, n_2, .., n_k, ...\rangle \tag{13.2.1}$$

to be the state where $|\psi^{(1)}\rangle$ is occupied by n_1 particles, $|\psi^{(2)}\rangle$ by n_2 particles, etc...

In other words, for bosons we have that:

$$|n_1, n_2, .., n_k, ...\rangle = \sqrt{\frac{N!}{n_1!n_2!...n_k!...}}\hat{S}_+\left(\bigotimes_{i=1}^{n_1}|\psi^{(1)}\rangle_i\right)\otimes\left(\bigotimes_{i=1}^{n_2}|\psi^{(2)}\rangle_{n_1+i}\right)...\left(\bigotimes_{i=1}^{n_k}|\psi^{(k)}\rangle_{...}\right)... \tag{13.2.2}$$

[1]the situation is more intricate for bosons where a state may be occupied by more than one particle

where $N = \sum_i n_i$, while for fermions:

$$|n_1, n_2, .., n_k, ...\rangle = \sqrt{N!}\hat{S}_-\left(\bigotimes_{i=1}^{n_1} |\psi^{(1)}\rangle_i\right) \otimes \left(\bigotimes_{i=1}^{n_2} |\psi^{(2)}\rangle_{n_1+i}\right)...\left(\bigotimes_{i=1}^{n_k} |\psi^{(k)}\rangle_{...}\right)... \tag{13.2.3}$$

where $n_i = 0, 1$ by the Pauli exclusion principle. The occupation representation is much more abstract and harder to use for fermions due to their state's anti-symmetry. Indeed, note that:

$$\begin{aligned} |..., n_i = 1, ..., n_j = 1, ...\rangle &= \sqrt{N!}\hat{S}_-(... \otimes |\psi^{(i)}\rangle \otimes ... \otimes |\psi^{(j)}\rangle ...) \\ \implies |..., n_j = 1, ..., n_i = 1, ...\rangle &= \sqrt{N!}\hat{S}_-(... \otimes |\psi^{(j)}\rangle \otimes ... \otimes |\psi^{(i)}\rangle ...) \\ &= -|..., n_i = 1, ..., n_j = 1, ...\rangle \end{aligned}$$

Clearly, the order in which we state the occupation of states is important, even though we're still denoting the same physical state.

States in the occupation representation constructed from a single-particle space $\mathcal{H}$ belong to the combined space of all possible states for an N-particle system, which we denote as $\mathcal{F}_N$:

$$\mathcal{F}_N = \text{Span}\{|n_1, n_2, ...,\rangle : \sum_i n_i = N\} \tag{13.2.4}$$

For example, letting $N = 2$ and $\mathcal{H} = \{|\uparrow\rangle, |\downarrow\rangle\}$ then:

$$\mathcal{F}_2 = \Bigg\{ \underbrace{|\uparrow\rangle_1 \otimes |\uparrow\rangle_2, |\downarrow\rangle_1 \otimes |\downarrow\rangle_2, \frac{1}{\sqrt{2}}(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 + |\downarrow\rangle_1 \otimes |\uparrow\rangle_1),}_{\in \text{Sym}_2 \mathcal{H}} \tag{13.2.5}$$

$$\underbrace{\frac{1}{\sqrt{2}}(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_1)}_{\in \bigwedge^2 \mathcal{H}} \Bigg\} \tag{13.2.6}$$

The **Fock space** $\mathcal{F}$ is defined as the direct sum of all $\mathcal{F}_i$:

$$\mathcal{F} = \bigoplus_{i=0}^{n} \mathcal{F}_i \tag{13.2.7}$$

13.3 Creation and annihilation operators

Bosonic operators

There are two important maps between $\mathcal{F}_N$ and $\mathcal{F}_{N+1}$, known as the creation and annihilation operators. The **bosonic creation operator** is defined as:

$$\begin{aligned} a_i^\dagger :&\mathcal{F}_N \to \mathcal{F}_{N+1} && (13.3.1) \\ &|n_1, ..., n_i, ...\rangle \mapsto \sqrt{n_i + 1}\,|n_1, ..., n_i + 1, ...\rangle && (13.3.2) \end{aligned}$$

so that (restricting the fock state to the occupation of $|\psi^{(i)}\rangle$ only):

$$\left\langle n_i + 1 \,\middle|\, a_i^\dagger \,\middle|\, n_i \right\rangle = \sqrt{n_i + 1} \tag{13.3.3}$$

$$\iff \langle n_i \,|\, a_i \,|\, n_i + 1\rangle = \left\langle n_i + 1 \,\middle|\, a_i^\dagger \,\middle|\, n_i \right\rangle^* = \sqrt{n_i + 1} \tag{13.3.4}$$

$$\iff a_i \,|n_i + 1\rangle = \sqrt{n_i + 1}\,|n_i\rangle \tag{13.3.5}$$

In other words, we have that the hermitian conjugate of the creation operator, known as the **bosonic annihilation operator**, is defined as:

$$a_i : \mathcal{F}_{N+1} \to \mathcal{F}_N \tag{13.3.6}$$

$$|n_1, ..., n_i + 1, ...\rangle \mapsto \sqrt{n_i + 1}\,|n_1, ..., n_i, ...\rangle \tag{13.3.7}$$

These operators allow us to create or destroy particles in a specific state. One must be wary however, since destroying too many particles eventually leads to the destruction of the vacuum state $|0\rangle$, where each state is not occupied, giving zero as a result.

We find that if $i \neq j$:

$$a_i a_j^\dagger \,|n_i, n_j\rangle = \sqrt{n_j + 1}\sqrt{n_i}\,|n_i - 1, n_j + 1\rangle \tag{13.3.8}$$

$$a_j^\dagger a_i \,|n_i, n_j\rangle = \sqrt{n_i}\sqrt{n_j + 1}\,|n_i - 1, n_j + 1\rangle \tag{13.3.9}$$

$$\iff [a_i, a_j^\dagger] = 0,\ i \neq j \tag{13.3.10}$$

while if $i = j$:

$$a_i a_i^\dagger \,|n_i\rangle = \sqrt{n_i + 1}\sqrt{n_i + 1}\,|n_i\rangle \tag{13.3.11}$$

$$a_i^\dagger a_i \,|n_i\rangle = \sqrt{n_1}\sqrt{n_i}\,|n_i\rangle \tag{13.3.12}$$

$$\iff [a_i, a_i^\dagger] = 1 \tag{13.3.13}$$

implying that:

$$[a_i, a_j^\dagger] = \delta_{ij} \tag{13.3.14}$$

Similarly, one finds that:

$$a_i^\dagger a_j^\dagger \,|n_i, n_j\rangle = \sqrt{n_j + 1}\sqrt{n_i + 1}\,|n_i + 1, n_j + 1\rangle \tag{13.3.15}$$

$$a_j^\dagger a_i^\dagger \,|n_i, n_j\rangle = \sqrt{n_i + 1}\sqrt{n_j + 1}\,|n_i + 1, n_j + 1\rangle \tag{13.3.16}$$

$$\iff [a_i^\dagger, a_j^\dagger] = 0,\ i \neq j \tag{13.3.17}$$

and since $[a_i^\dagger, a_i^\dagger] = 0$, we find that:

$$[a_i^\dagger, a_j^\dagger] = 0 \tag{13.3.18}$$

Therefore:

$$[a_i^\dagger, a_j^\dagger]^\dagger = [a_j, a_i] = 0 \tag{13.3.19}$$

giving:

$$[a_i, a_j] = 0 \tag{13.3.20}$$

These relations define the commutator algebra for bosonic creation/annihilation operators. Moreover, we may also use these operators to generate the Fock space from the vacuum state $|0\rangle$, since:

$$|n_1, n_2, ..., n_i, ...\rangle = \frac{1}{\sqrt{n_1! n_2! ... n_i! ...}} \prod_{i=1}^{N} (a_i^\dagger)^{n_i} |0\rangle \tag{13.3.21}$$

Finally, (13.3.12) suggests that we define a new operator, the **occupation number operator** $\hat{n}_i$, as the following automorphism:

$$\begin{aligned} \hat{n}_i : &\mathcal{F}_N \to \mathcal{F}_N &\quad (13.3.22)\\ &|n_1, ..., n_i, ...\rangle \mapsto n_i |n_1, ..., n_i, ...\rangle &\quad (13.3.23)\end{aligned}$$

which gives the occupation number of the ith state.

Fermionic operators

Just as in the case of bose statistics, we may define creation and annihilation operators for fermi statistics. However, care must be taken due to the exchange anti-symmetry of fermions, and a necessary revision to the bosonic operator definition will therefore be required.

The **fermionic creation operator** is defined as:

$$\begin{aligned} c_i^\dagger : &\mathcal{F}_N \to \mathcal{F}_{N+1} &\quad (13.3.24)\\ &|n_1, ..., n_i, ...\rangle \mapsto (-1)^{s_i} \sqrt{n_i + 1} |n_1, ..., n_i + 1, ...\rangle &\quad (13.3.25)\end{aligned}$$

where $s_i = \sum_{k=1}^{n_i - 1} n_k$.

Consequently, we see that (restricting the fock state to the occupation of $|\psi^{(i)}\rangle$ only):

$$\begin{aligned} &\left\langle n_i + 1 \middle| c_i^\dagger \middle| n_i \right\rangle = (-1)^{s_i} &\quad (13.3.26)\\ \iff &\langle n_i | a_i | n_i + 1\rangle = \left\langle n_i + 1 \middle| c_i^\dagger \middle| n_i \right\rangle^* = (-1)^{s_i} &\quad (13.3.27)\\ \iff &c_i |n_i + 1\rangle = (-1)^{s_i} |n_i\rangle &\quad (13.3.28)\end{aligned}$$

In other words, we have that the hermitian conjugate of the creation operator, known as the **fermionic annihilation operator**, is defined as:

$$\begin{aligned} c_i : &\mathcal{F}_{N+1} \to \mathcal{F}_N &\quad (13.3.29)\\ &|n_1, ..., n_i + 1, ...\rangle \mapsto (-1)^{s_i} \sqrt{n_i + 1} |n_1, ..., n_i, ...\rangle &\quad (13.3.30)\end{aligned}$$

To understand the significance of the $(-1)^{s_i}$ term, consider:

$$c_j \,|\underbrace{n_i = 1, ..., n_k = 1}_{s_i}, n_j = 1\rangle = -c_j \,|n_j = 1, n_{i+1} = 1, ..., n_k = 1, n_i = 1\rangle \tag{13.3.31}$$

$$= -\,|\underbrace{n_{i+1} = 1, ..., n_k = 1}_{s_i - 1}, n_i = 1\rangle \tag{13.3.32}$$

$$= (-1)(-1)^{s_i - 1} \,|n_i = 1, n_{i+1} = 1, ..., n_k = 1\rangle \tag{13.3.33}$$

$$= (-1)^{s_i} \,|n_i = 1, n_{i+1} = 1, ..., n_k = 1\rangle \tag{13.3.34}$$

and similarly:

$$c_j \,|\underbrace{n_i = 1, ..., n_k = 1}_{s_i}\rangle = |n_j = 1, n_i = 1, ..., n_k = 1\rangle \tag{13.3.35}$$

$$= (-1)^{s_i} \,|n_i = 1, ..., n_k = 1, n_j = 1\rangle \tag{13.3.36}$$

We see that the definition of the fermionic creation and annihilation operators still have the action of creating and annihilating fermions, but now taking exchange degeneracy into account.

Therefore:

$$|n_i = 1, n_j = 0\rangle = |n_j = 0, n_i = 1\rangle \text{ , and } |n_i = 1, n_j = 1\rangle = -\,|n_j = 1, n_i = 1\rangle \tag{13.3.37}$$

so that:

$$c_i c_j^\dagger \,|n_i = 1\rangle = c_i \,|n_j = 1, n_i = 1\rangle = c_i(-\,|n_i = 1, n_j = 1\rangle) = -\,|n_j = 1\rangle \tag{13.3.38}$$

which agrees with our definition of c_i since $s_j = 1$ and $s_i = 0$ gives a sign change. Similarly, we have that:

$$c_i c_j \,|n_i = 1, n_k = 1, n_j = 1\rangle = c_i(-\,|n_k = 1, n_i = 1\rangle) = c_i(|n_i = 1, n_k = 1\rangle) = |n_k = 1\rangle \tag{13.3.39}$$

which agrees with our definition of c_i since $s_j = 2$ and $s_i = 0$ give no sign changes. We can use these results (and similar ones) to evaluate the commutation relations for fermionic operators.

We find that if $i \neq j$:

$$c_i c_j^\dagger \,|n_i = 1, n_j = 1\rangle = c_i c_j^\dagger \,|n_j = 1\rangle = c_i c_j^\dagger \,|0\rangle = 0 \tag{13.3.40}$$

$$c_j^\dagger c_i \,|n_i = 1, n_j = 1\rangle = c_j^\dagger c_i \,|n_j = 1\rangle = c_j^\dagger c_i \,|0\rangle = 0 \tag{13.3.41}$$

and:

$$c_i c_j^\dagger \,|n_i = 1\rangle = c_i \,|n_j = 1, n_i = 1\rangle = -\,|n_i = 0, n_j = 1\rangle \tag{13.3.42}$$

$$c_j^\dagger c_i \,|n_i = 1\rangle = c_j^\dagger \,|0\rangle = |n_j = 1\rangle \tag{13.3.43}$$

while if $i = j$:

$$c_i c_i^\dagger |n_i = 1\rangle = 0, \qquad c_i c_i^\dagger |0\rangle = |0\rangle \tag{13.3.44}$$

$$c_i^\dagger c_i |n_i = 1\rangle = |n_i = 1\rangle\,,\ c_i^\dagger c_i |0\rangle = 0 \tag{13.3.45}$$

$$\Longleftrightarrow \{c_i, c_i^\dagger\} = 1 \tag{13.3.46}$$

implying that:

$$\{c_i, c_j^\dagger\} = \delta_{ij} \tag{13.3.47}$$

Similarly, one finds that the only non-zero effect of $c_i^\dagger c_j^\dagger$ is on the vacuum:

$$c_i^\dagger c_j^\dagger |0\rangle = |n_i = 1, n_j = 1\rangle \tag{13.3.48}$$

$$c_j^\dagger c_i^\dagger |0\rangle = |n_j = 1, n_i = 1\rangle = -\,|n_i = 1, n_j = 1\rangle \tag{13.3.49}$$

$$\Longleftrightarrow \{c_i^\dagger, c_j^\dagger\} = 0,\ i \neq j \tag{13.3.50}$$

and since $\{c_i^\dagger, c_i^\dagger\} = 0$, we find that:

$$\{c_i^\dagger, c_j^\dagger\} = 0 \tag{13.3.51}$$

Therefore:

$$\{c_i^\dagger, c_j^\dagger\}^\dagger = \{c_j, c_i\} = 0 \tag{13.3.52}$$

giving:

$$\{c_i, c_j\} = 0 \tag{13.3.53}$$

These are the anti-commutation relations for fermionic creation/annihilation operators, and are equivalent to the bosonic relations if we replace the anti-commutator by a commutator. Moreover, we may again use these operators to generate the Fock space from the vacuum state $|0\rangle$, since:

$$|n_1, n_2, ..., n_i, ...\rangle = \frac{1}{\sqrt{n_1! n_2! ... n_i! ...}} \prod_{i=1}^{N} (c_i^\dagger)^{n_i} |0\rangle \tag{13.3.54}$$

where the ordering of the products is as follows:

$$\prod_{i=1}^{N} (c_i^\dagger)^{n_i} = (c_1^\dagger)^{n_1} (c_2^\dagger)^{n_2} ... \tag{13.3.55}$$

Finally, the occupation number operator $\hat{n}_i$ is defined as usual, only that now its eigenvalue spectrum is restricted to 0 and 1, due to the Pauli exclusion principle.

General summary

In summary, if we define the following generalized commutator:

$$[\hat{A}, \hat{B}]_\eta = \hat{A}\hat{B} - \eta \hat{B}\hat{A} \tag{13.3.56}$$

then the generalized creation/annihilation operators $a_i^\dagger, a_i$ satisfy the following algebra:

$$[a_i, a_j^\dagger]_\eta = \delta_{ij},\ [a_i, a_j]_\eta = [a_i^\dagger, a_j^\dagger]_\eta = 0 \tag{13.3.57}$$

13.4 Field operators

The creation and annihilation operators may be used to convert operators in first quantization into **field operators** in second quantization.

In vague terms, a field operator is a field which assigns an operator to every point in real space. If we let $\{|\psi_i\rangle\}$ be a basis of a Hilbert space equipped with the continuous position basis $\{|\mathbf{r}\rangle\}$, then we define:

$$\Psi^\dagger(\mathbf{r}) = \sum_i \psi_i^*(\mathbf{r}) a_i^\dagger,\ \Psi(\mathbf{r}) = \sum_i \psi_i(\mathbf{r}) a_i \tag{13.4.1}$$

As *Bruus and Flensberg* [**?**] puts it, these field operators are the linear combination of "all possible ways to add a particle to the system at $\mathbf{r}$". An important special case of (13.4.1) is when we use the momentum basis $|\mathbf{k}\rangle = |\psi_i\rangle$ normalized over some volume $\mathcal{V}$. Then we find that:

$$\Psi^\dagger(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}} a_{\mathbf{k}}^\dagger,\ \Psi(\mathbf{r}) = \frac{1}{\sqrt{\mathcal{V}}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} a_{\mathbf{k}} \tag{13.4.2}$$

$$\iff a_{\mathbf{k}}^\dagger = \int e^{i\mathbf{k}\cdot\mathbf{r}} \Psi^\dagger(\mathbf{r})\, d\mathbf{r},\ a_{\mathbf{k}} = \int e^{-i\mathbf{k}\cdot\mathbf{r}} \Psi(\mathbf{r})\, d\mathbf{r} \tag{13.4.3}$$

where we used the fact that:

$$\int e^{-i(\mathbf{k}-\mathbf{q})\cdot\mathbf{r}} d\mathbf{r} = \mathcal{V}\delta_{\mathbf{k}\mathbf{q}} \tag{13.4.4}$$

Clearly, these represent Fourier transform relations between the creation/annihilation operators and the field operators.

It is easy to see that:

$$[\Psi(\mathbf{r}_1), \Psi^\dagger(\mathbf{r}_2)]_\eta = \frac{1}{\mathcal{V}}\left[\sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}_1} a_{\mathbf{k}}, \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{r}_2} a_{\mathbf{q}}^\dagger\right]_\eta \tag{13.4.5}$$

$$= \frac{1}{\mathcal{V}} \sum_{\mathbf{k}\mathbf{q}} e^{i(\mathbf{k}\cdot\mathbf{r}_1 - \mathbf{q}\cdot\mathbf{r}_2)} [a_{\mathbf{k}}, a_{\mathbf{q}}^\dagger]_\eta \tag{13.4.6}$$

$$= \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}_2 - \mathbf{r}_1)} \tag{13.4.7}$$

$$= \delta(\mathbf{r}_2 - \mathbf{r}_1) \tag{13.4.8}$$

and similarly

$$[\Psi(\mathbf{r}_1), \Psi(\mathbf{r}_2)]_\eta = [\Psi^\dagger(\mathbf{r}_1), \Psi^\dagger(\mathbf{r}_2)]_\eta = 0 \tag{13.4.9}$$

Representing single-body operators

Consider a single particle operator $\hat{f}$ acting on $\mathcal{H}$. In the full product space $\mathcal{H}^N$, then we would define:

$$\hat{f} = \mathbb{1} \otimes \mathbb{1} \otimes ... \otimes \hat{f} \otimes \mathbb{1}... \tag{13.4.10}$$

to be the single particle operator acting on the ith particle Hilbert space. Taking the sum over all particles, we recover the one-body operator:

$$\hat{F} = \sum_i \hat{f}_i \tag{13.4.11}$$

which in a $\{|i\rangle\}$ basis of $\mathcal{H}$ reads:

$$\hat{F} = \sum_{k,l} f_{kl} \sum_q |k\rangle_q \langle l|_q \, , \; f_{kl} = \left\langle k \,\middle|\, \hat{f} \,\middle|\, l \right\rangle \tag{13.4.12}$$

Our goal is to second quantize the expression $\sum_q |k\rangle_q \langle l|_q$, and do so by investigating its effect on some fock state $|n_i, n_j, ...\rangle$. We find that in first quantization:

$$\sum_q |k\rangle_q \langle l|_q \sqrt{\frac{N!}{n_i! n_j! ...}} \hat{S}_\pm \left(\bigotimes_{m=1}^{n_i} |i\rangle_m \right) \otimes \left(\bigotimes_{m=1}^{n_j} |j\rangle_{n_i+m} \right) ... \tag{13.4.13}$$

$$= \sqrt{\frac{N!}{n_i! n_j! ...}} \hat{S}_\pm \sum_q |k\rangle_q \langle l|_q \left(\bigotimes_{m=1}^{n_j} |i\rangle_m \right) \otimes \left(\bigotimes_{m=1}^{n_2} |j\rangle_{n_i+m} \right) ... \tag{13.4.14}$$

since $\hat{F}$ is exchange invariant, and therefore commutes with $\hat{S}_\pm$.

We can expand the sum in q to find that (we omit $\otimes$ to save space):

$$q = 1 : |k\rangle_1 \langle l \mid i\rangle_1 |i\rangle_2 ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |j\rangle_{n_i+n_j} ... |u\rangle_q ... \tag{13.4.15}$$

$$q = 2 : + |i\rangle_1 |k\rangle_2 \langle l \mid i\rangle_2 |i_3\rangle ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |j\rangle_{n_i+n_j} ... |u\rangle_q ... \tag{13.4.16}$$

$$+ \ldots \tag{13.4.17}$$

$$q : + |i\rangle_1 |i\rangle_2 ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |k\rangle_q \langle l \mid u\rangle_q ... \tag{13.4.18}$$

$$+ \ldots \tag{13.4.19}$$

In the qth line, we will get that $|u\rangle_q \to \delta_{lu} |k\rangle_q$, where $|u\rangle_q$ is whatever state is in the qth position. Consequently, the only lines that will survive will be the ones with state $|l\rangle_q$ in the appropriate position q. Since there will be n_l particles in the state $|l\rangle$, this will lead to n_l lines not vanishing. Each of these lines will also be some permutation of $|i\rangle_1 |i\rangle_2 ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |k\rangle_q ...$, and since $\hat{S}_\pm$ commutes with $\hat{P}_\sigma$ for any $\sigma \in S_N$, we find that:

$$\sum_q |k\rangle_q \langle l|_q \sqrt{\frac{N!}{n_i! n_j! ...}} \hat{S}_\pm \left(\bigotimes_{m=1}^{n_i} |i\rangle_m \right) \otimes \left(\bigotimes_{m=1}^{n_j} |j\rangle_{n_i+m} \right) ... \tag{13.4.20}$$

$$= n_l \sqrt{\frac{N!}{n_i! n_j! ...}} \hat{S}_\pm |i\rangle_1 |i\rangle_2 ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |k\rangle_q ... \tag{13.4.21}$$

and since:

$$|n_i, ..., n_l - 1, ..., n_k + 1, ...\rangle = \sqrt{\frac{N!}{n_i!...(n_l-1)!...(n_k+1)!...}}\hat{S}_\pm |i\rangle_1 |i\rangle_2 ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |k\rangle_q ... \tag{13.4.22}$$

we find that:

$$\sum_q |k\rangle_q \langle l|_q |n_i, ..., n_l, ..., n_k, ...\rangle \tag{13.4.23}$$

$$= n_l \sqrt{\frac{N!}{n_i!n_j!...}}\hat{S}_\pm |i\rangle_1 |i\rangle_2 ... |i\rangle_{n_i} |j\rangle_{n_i+1} ... |k\rangle_q ... \tag{13.4.24}$$

$$= n_l \sqrt{\frac{N!}{n_i!...n_l!...n_k!...}}\sqrt{\frac{n_i!...(n_l-1)!...(n_k+1)!...}{N!}} |n_i, ..., n_l - 1, ..., n_k + 1, ...\rangle \tag{13.4.25}$$

$$= \sqrt{n_l}\sqrt{n_k+1} |n_i, ..., n_l - 1, ..., n_k + 1, ...\rangle \tag{13.4.26}$$

$$= a_k^\dagger a_l |n_i, ..., n_l, ..., n_k, ...\rangle \tag{13.4.27}$$

Finally, we get the very elegant representation of a one-body operator:

$$\hat{F} = \sum_{kl} f_{kl} \hat{a}_k^\dagger \hat{a}_l \tag{13.4.28}$$

If we are working in a Hilbert space embedded with a position representation then we may also write that:

$$\hat{F} = \sum_{kl} f_{kl} \hat{a}_k^\dagger \hat{a}_l = \sum_{kl} \int \psi_k^*(\mathbf{r}) \hat{f} \psi_l(\mathbf{r}) \, d\mathbf{r} \hat{a}_k^\dagger \hat{a}_l = \int \Psi_k^*(\mathbf{r}) \hat{f} \Psi_l(\mathbf{r}) \, d\mathbf{r} \tag{13.4.29}$$

Representing two-body operators

We begin by deriving a useful property of creation/annihilation operators. Firstly note that the commutator algebra for these operators may be written as:

$$a_k a_j^\dagger = \eta a_j^\dagger a_k + \delta_{jk}, \; a_k a_l = \eta a_l a_k \tag{13.4.30}$$

where $\eta = 1$ for bosons and $\eta = -1$ for fermions. Then:

$$a_i^\dagger a_k a_j^\dagger a_l = a_i^\dagger (\eta a_j^\dagger a_k + \delta_{jk}) a_l \tag{13.4.31}$$

$$= \eta a_i^\dagger a_j^\dagger a_k a_l + \delta_{jk} a_i^\dagger a_l \tag{13.4.32}$$

$$= \eta^2 a_i^\dagger a_j^\dagger a_l a_k + \delta_{jk} a_i^\dagger a_l \tag{13.4.33}$$

$$= a_i^\dagger a_j^\dagger a_l a_k + \delta_{jk} a_i^\dagger a_l \tag{13.4.34}$$

Now consider a two-body operator written as $\hat{g}_{qq'} = \hat{f}_q \hat{h}_q'$ where $\hat{f}_q$ acts on $\mathcal{H}_q$ and $\hat{g}_{q'}$ acts on $\mathcal{H}_{q'}$. Then, we find that the total two-body operator may be written as:

$$\hat{G} = \frac{1}{2} \sum_{q \neq q'} \hat{g}_{qq'} \tag{13.4.35}$$

where $\frac{1}{2}$ takes care of double counting, and we discard $q = q'$ terms since a two-body operator must involve two different particles.

Therefore:

$$\hat{G} = \frac{1}{2}\sum_{q\neq q'} \hat{g}_{qq'} = \frac{1}{2}\Big(\sum_q \hat{f}_q \sum_{q'} \hat{g}_{q'} - \sum_q \hat{f}_q\hat{g}_q\Big) \tag{13.4.36}$$

$$= \frac{1}{2}\Big(\hat{F}\hat{G} - \sum_q \hat{f}_q\hat{g}_q\Big) \tag{13.4.37}$$

$$\tag{13.4.38}$$

Now we use the fact that $\hat{F} = \sum_q \hat{f}_q$, $\hat{G} = \sum_{q'} \hat{f}_{q'}$ and $\sum_q \hat{f}_q\hat{g}_q$ are single-body operators, and thus have a field representation of the type in (13.4.28):

$$\hat{G} = \frac{1}{2}\Big(\sum_{ik} f_{ik} a_i^\dagger a_k \sum_{jl} g_{jl} a_j^\dagger a_l - \sum_{il} (fh)_{il} a_i^\dagger a_l\Big) \tag{13.4.39}$$

$$= \frac{1}{2}\Big(\sum_{ijkl} f_{ik} g_{jl} a_i^\dagger a_k a_j^\dagger a_l - \sum_{il} (fh)_{il} a_i^\dagger a_l\Big) \tag{13.4.40}$$

$$= \frac{1}{2}\Big(\sum_{ijkl} f_{ik} g_{jl} a_i^\dagger a_j^\dagger a_l a_k + \sum_{ikjl} f_{ik} g_{jl} \delta_{jk} a_i^\dagger a_l - \sum_{il} (fh)_{il} a_i^\dagger a_l\Big) \tag{13.4.41}$$

$$= \frac{1}{2}\Big(\sum_{ijkl} f_{ik} g_{jl} a_i^\dagger a_j^\dagger a_l a_k + \sum_{ijl} f_{ij} g_{jl} a_i^\dagger a_l - \sum_{ijl} f_{ij} h_{jl} a_i^\dagger a_l\Big) \tag{13.4.42}$$

$$= \frac{1}{2}\sum_{ijkl} f_{ik} g_{jl} a_i^\dagger a_j^\dagger a_l a_k \tag{13.4.43}$$

Note that the matrix elements of $\hat{g}, \hat{f}, \hat{h}$ are related by:

$$g_{ijkl} = \langle i|_q \langle j|_{q'} |\hat{g}| |k\rangle_q |l\rangle_{q'} = \langle i|_{q'} \langle j|_q |\hat{f}_q \hat{h}_q'| |k\rangle_q |l\rangle_{q'} = f_{ik} g_{jl} \tag{13.4.44}$$

so that:

$$\hat{G} = \frac{1}{2}\sum_{ijkl} g_{ijkl} a_i^\dagger a_j^\dagger a_l a_k \tag{13.4.45}$$

Luckily, any two-body operator may be expanded as a power series in one-particle operators:

$$G = \sum_{\alpha\beta} c_{\alpha\beta} \sum_{q\neq q'} \hat{f}_q^\alpha \hat{h}_{q'}^\beta \tag{13.4.46}$$

$$= \frac{1}{2}\sum_{ikjl} (f^\alpha)_{ik} (g^\beta)_{jl} a_i^\dagger a_j^\dagger a_l a_k \tag{13.4.47}$$

$$= \frac{1}{2}\sum_{ikjl} g_{ijkl} a_i^\dagger a_j^\dagger a_l a_k \tag{13.4.48}$$

Change of basis

Finally, we must comment on how changes of basis affect the field representations we have derived. We have already observed that the change from the position to the momentum basis is given by a fourier transform.

More generally, we have that given two bases $\{|u_i\rangle\}$ and $\{|v_i\rangle\}$ of $\mathcal{H}$. Then, for any $|\psi\rangle \in \mathcal{H}$:

$$\hat{a}^\dagger_{u_i} |0\rangle = |u_i\rangle = \sum_j \langle v_j \,|\, u_i\rangle |v_j\rangle = \sum_j \langle v_j \,|\, u_i\rangle \hat{a}^\dagger_{v_j} |0\rangle \tag{13.4.49}$$

implying that:

$$a^\dagger_{u_i} = \sum_j \langle v_j \,|\, u_i\rangle a^\dagger_{v_j} \implies a_{u_i} = \sum_j \langle u_i \,|\, v_j\rangle a_{v_j} \tag{13.4.50}$$

Clearly, we see that using $\{|u_i\rangle\} = \{|\mathbf{r}\rangle\}$ then $\hat{a}_\mathbf{r} = \sum_j \langle \mathbf{r} \,|\, v_j\rangle a_{v_j}$ which is just the field operator $\Psi(\mathbf{r})$ we defined earlier.

Using a change of basis allows us to derive in a much simpler way the field representation of diagonalizable operators. Indeed, suppose we have some one-body operator $\hat{f}$ with eigenbasis $\{|\psi_i\rangle\}$ and eigenvectors λ_i. Then:

$$\hat{F} = \sum_i \lambda_i \hat{n}_i = \sum_i \lambda_i a^\dagger_{\psi_i} a_{\psi_i} \tag{13.4.51}$$

Consequently, using another basis $\{|\phi_j\rangle\}$ then

$$\hat{F} = \sum_i \lambda_i a^\dagger_{\psi_i} a_{\psi_i} = \sum_i \lambda_i \sum_k \langle \phi_k \,|\, \psi_i\rangle a^\dagger_{\phi_k} \sum_j \langle \psi_i \,|\, \phi_j\rangle a_{\phi_j} \tag{13.4.52}$$

$$= \sum_{ikj} \langle \phi_k \,|\, \psi_i\rangle \left\langle \psi_i \,\middle|\, \hat{f} \,\middle|\, \psi_i \right\rangle \langle \psi_i \,|\, \phi_j\rangle a^\dagger_{\phi_k} a_{\phi_j} \tag{13.4.53}$$

$$= \sum_{kj} \left\langle \phi_k \,\middle|\, \hat{f} \,\middle|\, \phi_j \right\rangle a^\dagger_{\phi_k} a_{\phi_j} \tag{13.4.54}$$

$$= \sum_{kj} f_{kj} a^\dagger_{\phi_k} a_{\phi_j} \tag{13.4.55}$$

as we found earlier. Similar arguments may be used to show that for two-body operators $\hat{G}$:

$$\hat{G} = \frac{1}{2} \sum_{ikjl} g_{ijkl} a^\dagger_i a^\dagger_j a_l a_k \tag{13.4.56}$$

Part III

Solid state physics

Solids: Boltzmann vs Einstein vs Debye

14

14.1 The heat capacity of solids

Recall from Thermodynamics that

$$C_p - C_V = \frac{VT\alpha^2}{\kappa_T}, \ \alpha = \frac{1}{V}\frac{\partial p}{\partial T}\bigg|_V, \ \kappa_T = -\frac{1}{V}\frac{\partial V}{\partial p}\bigg|_T \tag{14.1.1}$$

where α is the thermal expansion coefficient and κ_T is the isothermal compressibility. For solids, the expansion coefficient is very small, so usually we can set the isobaric and isochoric heat capacities to be equal to each other:

$$C_p \approx C_V = C \tag{14.1.2}$$

The goal of this chapter will be to study the heat capacity using different models, namely the Boltzmann, Einstein and Debye models.

14.2 Boltzmann model

The simplest model for a solid was given in 1896 by Boltzmann, where he took each atom in a solid to reside in a (classical) harmonic potential. The resulting heat capacity can be calculated using the typical method of writing down the partition function, finding the free energy and then using the appropriate thermodynamic relation (see statistical physics lecture notes). There is however a much faster way to calculate C.

The short-cut is to use the equipartition theorem. We argue that there are six degrees of freedom in our system, p_x, p_y, p_z (due to kinetic energy) and x, y, z (due to the harmonic well being translationally invariant). Since each degree of freedom contributes $\frac{1}{2}Nk_B$ to the heat capacity we should expect

$$C = 3Nk_B \tag{14.2.1}$$

which is precisely what the rigorous method yields.

The law in (14.2.1) is known as the **Dulong-Petit law**. While the Boltzmann model works fairly well at high temperatures where the classical picture holds, we get deviations at lower temperatures for certain materials, most notably for Diamond where $C < 3Nk_B$ even at room temperature.

14.3 Boltzmann model

In 1907 Einstein solved this inconsistency by adding some quantum flavour to Boltzmann's model (19 years before Schrodinger's paper even came out!). Instead of treating the harmonic potential as classical, we can use quantum mechanics and use the quantised energy levels of the quantum harmonic oscillator:

$$E_n = 3\hbar\omega\left(n + \frac{1}{2}\right) \tag{14.3.1}$$

From the statistical physics lecture notes we find that

$$U = 3\hbar\omega\left(n_B\beta\hbar\omega + \frac{1}{2}\right), \; n_B = \frac{1}{e^{\beta\hbar\omega} - 1} \tag{14.3.2}$$

where n_B is known as the **Bose factor**, and roughly represents the average excitation level of the system at a given temperature. Differentiating with respect to temperature we find that

$$C = 3Nk_B(\beta\hbar\omega)^2 \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2} \tag{14.3.3}$$

In the high temperature limit then $e^{\beta\hbar\omega} \approx 1 + \beta\hbar\omega$ so to first order

$$C \approx 3Nk_B(1 + \hbar\beta\omega) \tag{14.3.4}$$

which (to zeroth order) reduces to the Dulong-Petit law. In the low temperature limit, on the other hand, we find that

$$C = 3Nk_B(\beta\hbar\omega)^2 e^{-\beta\hbar\omega} \tag{14.3.5}$$

so we get an exponential temperature suppression factor.

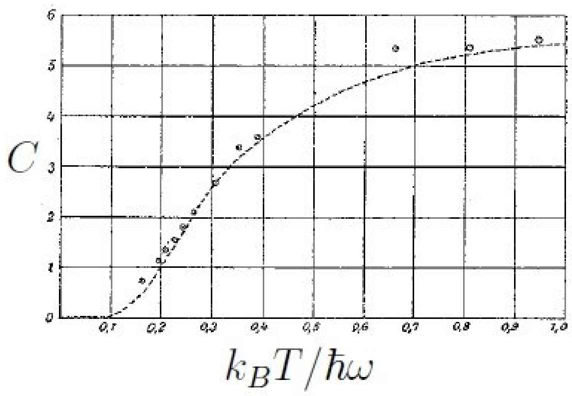


Figure 14.1. Plot of the heat capacity according to the Einstein model, from Einstein's original 1907 paper.

Unfortunately, Einstein's model still has some shortcomings, most notably at really low temperatures where the heat capacity seems to have a $\sim T^3$ type of behaviour.

14.4 Debye model

This inconsistency was solved by Debye in 1912 by noting that if we displace an electron from one state to another, it will push on the other electrons, which will themselves move and form sound waves, quantized modes known as phonons. Thus, instead of considering a set of N harmonic oscillators, let's consider a set of phonon waves with N possible modes (up to a polarisation factor).

The internal energy is now given by the sum of the energies for each mode ω_i:

$$U = \sum_i \hbar\omega_i \left(n_B(\beta\hbar\omega_i) + \frac{1}{2} \right) \tag{14.4.1}$$

where the sum is over all phonon modes i. We can have three different polarisations of our sound waves, and we can assume that the speed of sound is independent of direction and polarisation. We find that

$$\sum_i (\text{modes}) \to \frac{3V}{(2\pi)^3} \int d^3\mathbf{k} \tag{14.4.2}$$

and using the dispersion relation $\omega(k) = v_D k$ then

$$\sum_i (\text{modes}) \to \frac{3V}{(2\pi)^3} \int d^3\mathbf{k} \to \frac{3V}{2\pi^2 v^3} \int_0^\infty \omega^2 d\omega \tag{14.4.3}$$

We can therefore read off the density of states as

$$g(\omega) = \frac{3V}{2\pi^2 v^3}\omega^2 = N\frac{9\omega^2}{\omega_D^2},\ \omega_D = \sqrt{6\pi^2 \frac{N}{V} v^3} \tag{14.4.4}$$

where we defined the **Debye frequency**. Consequently the total energy reads

$$U = \int_0^\infty g(\omega)\hbar\omega \left(n_B(\beta\hbar\omega) + \frac{1}{2} \right) d\omega \tag{14.4.5}$$

This is slightly worrying, the $\frac{1}{2}$ factor will give a diverging term in the integrand, and we do not want infinities. However, this term will be temperature independent, and since we are interested in the heat capacity it will drop off when differentiating it with respect to temperature. Later we will see how to deal with this infinity more justifiably, so we proceed with resolve and ignore the factor for now:

$$U = \frac{9N\hbar}{\omega_D^3} \int_0^\infty \frac{\omega^3}{e^{\beta\hbar\omega} - 1} \tag{14.4.6}$$

and recognising the Bose-Einstein integral we get

$$U = \frac{3}{5} N \frac{(k_B T)^4}{(\hbar\omega_D)^2} \pi^4 \tag{14.4.7}$$

which when differentiated gives the Debye heat capacity

$$C = \frac{12}{5} N k_B \left(\frac{T}{T_D}\right) \pi^4, \; T_D = \frac{\hbar\omega_D}{k_B} \tag{14.4.8}$$

Here T_D is known as the **Debye temperature**. This result gives the required low temperature T^3 behaviour, but it seems like this came at the cost of losing the empirically verified high temperature result $C \approx 3Nk_B$. Going back to the Boltzmann model remember that the factor of 3 came from the fact that we have 3 degrees of freedom, but in Debye's model we included an infinite number of wave modes, and thus infinitely many degrees of freedom. However, it is incorrect to assume that arbitrarily high frequency sound waves can propagate through a solid since if the wavelength of this wave becomes smaller than the atomic spacing in the solid then the wave will not propagate at all. To fix this issue we can impose an ad-hoc high frequency cut-off ω_c so that

$$\int_0^{\omega_c} g(\omega)\, d\omega = 3N \tag{14.4.9}$$

This is saying that the total number of permissible wave modes must be equal to the number of degrees of freedom in the system. Intuitively this makes sense because to fully specify a system it suffices to know how many particles occupy each occupable mode, so that if there are $3N$ modes then we must specify $3N$ occupation numbers. Consequently we find that

$$\frac{3V}{2\pi^2 v^3} \int_0^{\omega_c} \omega^2 \, d\omega = \frac{V}{2\pi^2 v^3} \omega_c^3 = 3N \implies \omega_c = \omega_D \tag{14.4.10}$$

The Debye frequency thus represents the highest frequency mode that can travel through the solid, above which the density of states vanishes. If we redo our calculation we see that

$$U = \int_0^{\omega_D} g(\omega)\hbar\omega \left(n_B(\beta\hbar\omega) + \frac{1}{2}\right) d\omega = \frac{9N\hbar}{\omega_D^3} \int_0^{\omega_D} \frac{\omega^3}{e^{\beta\hbar\omega} - 1} \tag{14.4.11}$$

where the zero-point energy contribution not longer diverges, but can be omitted since it is temperature independent. We obtain the heat capacity by differentiating

$$C_V = \frac{9Nk_B\hbar^2}{\omega_D^3} \beta^2 \int_0^{\omega_D} \frac{\omega^4 e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2} \, d\omega \tag{14.4.12}$$

This can be rewritten more nicely as

$$C_V = 9Nk_B \left(\frac{T}{T_D}\right)^3 \int_0^{T_D/T} \frac{x^4 e^x}{(e^x - 1)^2} \, dx \tag{14.4.13}$$

Note that at low temperature it does not matter whether we cut off the integral at T_D/T or

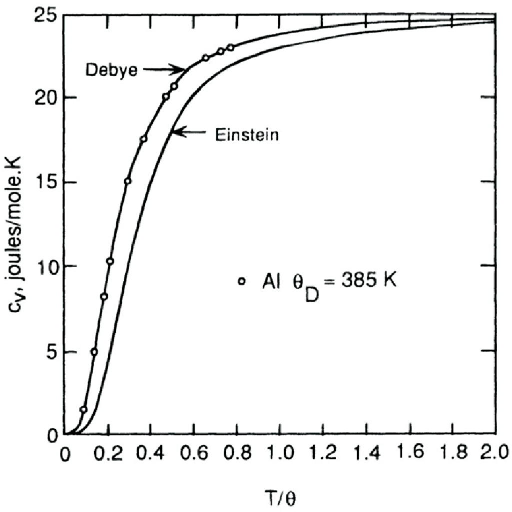


Figure 14.2. The heat capacity predicted by the Einstien model vs. the Debye model.

∞, so the result we previously found

$$C_V = \frac{12}{5} N k_B \left(\frac{T}{T_D}\right) \pi^4, \qquad \text{at low } T \tag{14.4.14}$$

still holds [1]. At high temperature instead we can Taylor expand the integrand

$$U = 9 N k_B \left(\frac{T}{T_D}\right)^3 \int_0^{T_D/T} x^2 \, dx = 3 N k_B \tag{14.4.16}$$

just as required by the Dulong-Petit law (and most importantly by experimental data). Other than the fact that the Debye model is just visibly a better fit to experimental data, its other strength is that unlike the previous models it has no fit parameter: ω_D, the Debye frequency, can be measured in a lab.

[1] it is easier to directly evaluate the infinite integral in U and differentiating than calculating the integral

$$\int_0^\infty \frac{x^4 e^x}{(e^x - 1)^2} = \frac{4\pi^4}{15} \tag{14.4.15}$$

Metals: Drude vs. Sommerfield

15

15.1 Drude model

Consider a system of electrons moving under the influence of a force $\mathbf{F}$, and which scatter with each other with scattering time τ. We assume that the scattering process is completely inelastic so that the momentum after a collision is zero (although this obviously does not hold for individual collisions, the average effect is that the particle can scatter in any direction and thus $\mathbf{p} = 0$).

If we picture a single electron with momentum $\mathbf{p}(t)$ at time t, then after some time dt we will find that

$$\mathbf{p}(t+dt) = \left(1 - \frac{dt}{\tau}\right)[\mathbf{p}(t) + \mathbf{F}dt] + \frac{dt}{\tau}\mathbf{0} \tag{15.1.1}$$

$$\implies \mathbf{p}(t+dt) - \mathbf{p}(t) = \left(\mathbf{F} - \frac{\mathbf{p}}{\tau}\right)dt \tag{15.1.2}$$

$$\implies \frac{d\mathbf{p}}{dt} = \mathbf{F} - \frac{\mathbf{p}}{\tau} \tag{15.1.3}$$

so we see that the scattering effects act as a drag force. If the electron is in an electromagnetic field then

$$\boxed{\frac{d\mathbf{p}}{dt} = -e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \frac{\mathbf{p}}{\tau}} \tag{15.1.4}$$

The steady state condition $\frac{d\mathbf{p}}{dt} = 0$ fixes the equilibrium **drift velocity**

$$-e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) - \frac{\mathbf{p}}{\tau} = 0 \tag{15.1.5}$$

Let's define the current density $\mathbf{j} = -ne\mathbf{v}$, then

$$\mathbf{E} = \frac{1}{ne}\mathbf{j} \times \mathbf{B} + \frac{m}{ne^2\tau}\mathbf{j} \tag{15.1.6}$$

so we get two contributions to the electric field, a contribution parallel to the current, and a **Hall electric field** perpendicular to the current.

Suppose initially that $\mathbf{B} = 0$, then we find that

$$\mathbf{j} = \sigma\mathbf{E}, \ \sigma = \frac{ne^2\tau}{m} \tag{15.1.7}$$

so the current produced is proportional to the applied electric field, with conductivity σ being the proportionality constant.

Now let's turn on the magnetic field: $\mathbf{B} \neq 0$. Suppose we measure the Hall electric field $\mathbf{E}_H$, then if we know n and e then we can evaluate the magnetic field via

$$\mathbf{E}_H = R_H \mathbf{B} \times \mathbf{j}, \; R_H = -\frac{1}{ne} \tag{15.1.8}$$

where R_H is known as the **Hall coefficient**. More specifically, assuming we have a constant magnetic field then we can let $\mathbf{B} = B\mathbf{z}$. This then yields

$$E_x = \frac{1}{ne} j_y B_z + \frac{m}{ne^2\tau} j_x \tag{15.1.9}$$

$$E_y = -\frac{1}{ne} j_x B_z + \frac{m}{ne^2\tau} j_x \tag{15.1.10}$$

$$E_z = \frac{m}{ne^2\tau} j_z \tag{15.1.11}$$

or in matrix notation

$$\mathbf{E} = \begin{pmatrix} \rho_\parallel & \rho_{xy} & 0 \\ -\rho_{xy} & \rho_\parallel & 0 \\ 0 & 0 & \rho_\parallel \end{pmatrix} \mathbf{j}, \; \rho_\parallel = \frac{m}{ne^2\tau}, \; \rho_{xy} = \frac{B_z}{ne^2} \tag{15.1.12}$$

We define the matrix $\boldsymbol{\rho}$ as the **resistivity**, and its inverse is the **conductivity**

$$\mathbf{j} = \begin{pmatrix} \sigma_\parallel & \sigma_{xy} & 0 \\ -\sigma_{xy} & \sigma_\parallel & 0 \\ 0 & 0 & \frac{1}{\rho_\parallel} \end{pmatrix} \mathbf{E}, \; \boxed{\sigma_{xy} = -\frac{\rho_{xy}}{\rho_\parallel^2 + \rho_{xy}^2}, \; \sigma_\parallel = \frac{\rho_\parallel}{\rho_\parallel^2 + \rho_{xy}^2}} \tag{15.1.13}$$

As expected we get a parallel resistivity $\rho_\parallel = \frac{m}{ne^2\tau} = \frac{1}{\sigma}$ which we calculated earlier, but now we also get off-diagonal resistivities given by ρ_{xy}, which correspond to longitudinal currents producing transverse Hall electric field. This electric field is proportional to B_z and inversely proportional to the carrier density n and charge e.

Unfortunately, the Drude model for conductivity is not always correct. Indeed, for materials such as Beryllium or Magesium the Hall resistance, which is supposed to be negative, actually takes positive values. Nevertheless, let's go on and calculate the thermal conductivity which we know from classical thermodynamics is given by:

$$\kappa = \frac{1}{3} n c_v \langle v \rangle \lambda \tag{15.1.14}$$

According to the Drude model the scattering length λ is $\lambda = \langle v \rangle \tau$, and $\langle v \rangle = \sqrt{\frac{8k_BT}{\pi m}}$ to find that

$$\kappa \overset{?}{=} \frac{4}{\pi} \frac{n\tau k_B^2 T}{m} \tag{15.1.15}$$

The question mark arises from dubiously inserting $c_v = \frac{3}{2}k_B$. a result which applies to ideal gases. We still do not know what τ is so it is difficult to compare with experiment. However, since $\sigma_\parallel = \frac{ne^2\tau}{m}$ we can take the ratio and remove τ to find the **Lorentz number**

L_R

$$L_R = \frac{\kappa}{T\sigma_{\|}} = \frac{4}{\pi}\left(\frac{k_B}{e}\right)^2 \approx 1.0 \times 10^{-8}\ \mathrm{W}\,\Omega\,\mathrm{K}^{-2} \tag{15.1.16}$$

Empirically the Lorentz number is 2.2×10^{-8} W Ω K^{-2}, a result known as the **Wiedemann-Franz law**.

Note that the Lorentz number is universal and should be the same for any metal, it only depends on fundamental quantities such as the Boltzmann constant and the electron charge. This result is quite intuitive since we should expect thermal and electrical conductivity to be governed by analogous transport processes, so taking the ratio of analogous transport coefficients should give a constant. Indeed we also find that the thermal current density $\mathbf{j}^q$ and the electrical current density $\mathbf{j}$ are related by $\mathbf{j}^q = \Pi\mathbf{j}$ where Π is the **Peltier coefficient**:

$$\Pi = -\frac{c_V T}{3e} \tag{15.1.17}$$

If we run current through a metal then this will produce a thermal current and cool one side of the material. However we cannot cool the material to arbitrarily low temperatures as the Joule heat dissipation I^2R can quickly dominate over the Peltier effect. We can get another universal constant from Π

$$S = \frac{\Pi}{T} = -\frac{c_V}{3e} \stackrel{?}{=} -\frac{k_B}{2e} \approx -0.4 \times 10^{-4}\ \mathrm{V}\,\mathrm{K}^{-1} \tag{15.1.18}$$

where in the last step we substituted $c_V = \frac{3}{2}k_B$. Experiments show that S should be about 10^{-2} times smaller! Of course, the mistake comes from haphazardly substituting $c_v = \frac{3}{2}k_B$, since the electrons in a metal are not ideal gas particles! In reality, $c_V \ll \frac{3}{2}k_B$, so S should be much larger. However this should also mean that L_R should be much smaller, how did we get such good agreement with experiment? The answer is that we made two cancelling errors, one was of course $c_v = \frac{3}{2}k_B$, but the other was $\langle v\rangle = \sqrt{\frac{8k_BT}{\pi m}}$ which again comes from the kinetic theory of ideal gases. In reality $\langle v\rangle$ is much larger, by about the same amount by which c_v is smaller than the ideal result. Inserting these two expressions into κ gave a roughly good approximation by pure coincidence, while for S this did not occur.

15.2 Sommerfield model

The reason we made the two cancelling errors in calculating the thermal conductivity was that we treated the electron gas as an ideal gas, while in reality it is a fermion gas. Recall that the average number of particles occupying a state with energy ϵ at temperature β

$$n(\epsilon) = \frac{1}{e^{\beta(\epsilon-\mu)} + 1} \tag{15.2.1}$$

and that $\mu(T=0) \equiv E_F$ is known as the Fermi energy which for a given particle density n takes the form

$$E_F = \frac{\hbar^2}{2m}(3\pi^2 n)^{2/3} = \frac{\hbar^2 k_F^2}{2m},\ k_F = (3\pi n)^{1/3} \tag{15.2.2}$$

In principle, one could calculate the average number of particles N as an integral, which would yield a Fermi-Dirac function. One could take a low temperature approximation of this expression and find the chemical potential as a function of n. Then one could similarly calculate the energy U as an integral and insert the expressions for μ to get something completely in terms of n and T. Finally one takes the derivative with respect to temperature to find C_V. This is the approach taken in the Statistical physics lecture notes. However, we can cheat a little and derive the same result with more physical intuition.

As we slightly raise the temperature above $T = 0$, the number of particles that move from the grey region to the white region is roughly the same. Consequently, to keep n fixed one does not need to change μ significantly, thus allowing us to approximate $\mu \approx E_F$. We also assume that

$$E(T) = E(T = 0) + \frac{\gamma}{2} N^* E^* \tag{15.2.3}$$

where N^* is the number of electrons that can be excited, and E^* is the average energy that each state can absorb. We also introduced $\frac{\gamma}{2}$ as a fudge factor. We see that the average number of electrons that can get excited (by thermal fluctuations) must be roughly within a distance $k_B T$ of E_F so

$$N^* = \int_{E_F + k_B T}^{E_F - k_B T} g(E) n_F(E) dE \approx g(E_F) k_B T \tag{15.2.4}$$

since $k_B T << E_F$. Moreover, by this argument the energy that each electron absorbs is roughly $k_B T$ so that

$$E(T) = E(T = 0) + \frac{\gamma}{2} g(E_F)(k_B T)^2 \tag{15.2.5}$$

Therefore, defining $T_F = \frac{E_F}{k_B}$ as the Fermi temperature and using $g(E_F) = \frac{3}{2}\frac{N}{E_F}$ then

$$C_V = \frac{\partial U}{\partial T} = \gamma k_B T g(E_F) = \boxed{\gamma \frac{3}{2} N k_B \frac{T}{T_F}} \tag{15.2.6}$$

where it turns out that $\gamma = \frac{\pi^2}{3}$.

15.3 Conclusions

The free-electron model is quite successful, it yielded good values for the heat capacity, thermal and electrical conductivity, the Peltier coefficient and so forth. Nevertheless, it still has several shortcoming which must be addressed. Firstly, experiments are able to measure the scattering time τ of electrons, and the resulting scattering length $\lambda = v_F \tau$ is unreasonable large, reaching values as high as 1mm at low temperatures. Furthermore, the anomalous sign of the Hall coefficient is still unaccounted for, and so are the various optical properties of metals. The main that we have been neglecting is the microscopic arrangement of the electrons in matter.

16 Vibration of solids

16.1 1D Monoatomic harmonic chain

The "old quantum theory" models we have constructed for electrons in solids and metals still have several shortcomings, such as the negativity of the Hall resistance and the impossibly large scattering length. One important feature of electrons in crystals that we have not exploited yet is the microscopic lattice symmetry. To investigate the fundamental aspects of electron arrangements in solids, we will investigate a mono-atomic harmonic chain.

Let's consider two atoms with separation x in an attractive potential $V(x)$ which we can expand about the equilibrium distance x_0 as a harmonic well:

$$V(x) = V(x^0) + \frac{\kappa}{2}(x - x^0)^2 + o((x - x^0)^3), \ \kappa = V'(x^0) \tag{16.1.1}$$

As long as we are at sufficiently low temperatures, this expansion is valid and the average separation between the atoms will be the equilibrium distance x^0 due to the symmetry of quadratic potentials. At high temperatures however, thermal fluctuations can cause the electrons to deviate outside of the quadratic well region. Since the potential is steeper on one side (small separation), this causes the average distance between the atoms to shift away from the equilibrium.

Consider an infinite chain of atoms with mass m separated by a distance a, known as the lattice constant, and connected with springs of constants κ. Let $x_n^0 = na$ be the nth atom's equilibrium position, and let $q_n = x_n - x_n^0$ be the deviation of atom n from its equilibrium. Newton's second law applied to the nth atom yields

$$m\ddot{q}_n = \kappa(q_{n+1} - q_n) + \kappa(q_{n-1} - q_n) \tag{16.1.2}$$

We make a wave ansatz $q_n = Ae^{i\omega t - ikx_n^0}$ and get

$$-Am\omega^2 e^{i\omega t - ikna} = \kappa(Ae^{i\omega t - ik(n+1)a} + Ae^{i\omega t - ik(n-1)a} - 2Ae^{i\omega t - ikna}) \tag{16.1.3}$$

$$\implies -m\omega^2 = \kappa(2 - 2\cos(ka)) \implies \boxed{\omega(k) = 2\sqrt{\frac{\kappa}{m}}\left|\sin\left(\frac{ka}{2}\right)\right|} \tag{16.1.4}$$

In the long wavelength, low wavenumber limit we should get sound waves, and indeed a

Taylor expansion quickly gives the correct dispersion relation

$$\omega = 2\sqrt{\frac{\kappa}{m}}\frac{|k|a}{2} = v|k|,\ v = \sqrt{\frac{\kappa}{m}}a \tag{16.1.5}$$

We also found that the speed of sound in this chain is given by the lattice spacing times the spring frequency. We could also obtain this result by first calculating the compressibility β

$$\beta = -\frac{1}{V}\frac{\partial V}{\partial p} = -\frac{1}{L}\frac{\partial L}{\partial F} = \frac{1}{\kappa a} \tag{16.1.6}$$

and then using the following result from fluid dynamics

$$v = \sqrt{\frac{1}{\rho\beta}} = \sqrt{\frac{1}{\frac{m}{a}\cdot\frac{1}{\kappa a}}} = \sqrt{\frac{\kappa}{m}}a \tag{16.1.7}$$

Note also that the dispersion relation has a maximum

$$\omega_{\text{max}} = 2\sqrt{\frac{\kappa}{m}} \text{ at } k = \frac{\pi}{a} \tag{16.1.8}$$

and the corresponding mode is given by $q_n(t) = Ae^{i\omega t}(-1)^n$ which corresponds to each atom being out of phase with its two neighbours.

16.2 Reciprocal space

Reciprocal space (or momentum space) is the space of k's, it is the Fourier transform of real space (or direct space). Interestingly, our modes are periodic in momentum space:

$$q_n(t) = Ae^{i\omega t - i(k+2\pi/a)na} = Ae^{i\omega t - ikna}e^{2\pi n} = Ae^{i\omega t - ikna} \tag{16.2.1}$$

Therefore, shifting our wavenumber by $\frac{2\pi}{a}$ we obtain the same physical mode, the periodicity of the lattice in real space yielded a periodicity in reciprocal space.

To understand the in the wavenumber, note that the following waves, with k differing by $2\pi/a$, predict the same value for the displacement of each mass. Since that is all the information we can obtain from the $q_n(t)$ functions, both solutions are equivalent.

We can then define the **reciprocal lattice** as the set of all points in momentum space equivalent to $k = 0$, they are given by

$$G_n = \frac{2\pi n}{a} \tag{16.2.2}$$

Note also that there is a fundamental interval in reciprocal space, the **Brillouin zone** (BZ), which contains all non-equivalent wavenumbers, it is the unit cell of the reciprocal lattice. Every point outside the BZ will be equivalent to some point inside the BZ. In our case the Brillouin zone is

$$BZ = \left(-\frac{\pi}{a}, \frac{\pi}{a}\right] \tag{16.2.3}$$

and since the spacing of the reciprocal lattice is $\frac{2\pi}{Na}$ we find that the total number of modes with periodic boundary conditions is

$$\text{number of modes } = \frac{2\pi/a}{2\pi/Na} = N \tag{16.2.4}$$

Exactly as Debye had predicted, there are as many normal modes as masses in the chain.

16.3 Quantum modes: phonons

We have already encountered phonons in the Debye model, they are quanta of vibrations which possess a quantized energy $\hbar\omega(\mathbf{k})$. Note that since two or more phonons can have the same mode, they follow Bose-Einstein statistics. The energy of a collection of phonons is then given by

$$U = \sum_{k\in BZ} \hbar\omega_k \left(\frac{1}{e^{\beta\hbar\omega_k} - 1} + \frac{1}{2} \right) \tag{16.3.1}$$

Assuming $a << L$ we can approximate the sum as an integral

$$\sum_{k\in BZ} \longrightarrow \frac{L}{2\pi} \int_{-\pi/a}^{\pi/a} dk \tag{16.3.2}$$

This is a good point to stop and review what dispersion relations the various quantum models use:

$$\text{Chain: } \omega(k) = 2\sqrt{\frac{\kappa}{m}} \left| \sin\left(\frac{ka}{2}\right) \right| \tag{16.3.3}$$

$$\text{Debye: } \omega(k) = v|k| \tag{16.3.4}$$

$$\text{Einstein: } \omega(k) = \omega_0 \tag{16.3.5}$$

Since the dispersion relation is even in k we can write

$$U = 2\int_0^{\pi/a} dk\ \hbar\omega(k) \left(\frac{1}{e^{\beta\hbar\omega(k)} - 1} + \frac{1}{2} \right) \tag{16.3.6}$$

Using the chain dispersion relation

$$d\omega = a\sqrt{\frac{\kappa}{m}} \cos\left(\frac{ka}{2}\right) dk = a\sqrt{\frac{\kappa}{m}} \sqrt{1-\omega^2} \tag{16.3.7}$$

we then write

$$U = 2a\int_0^{2\sqrt{\frac{\kappa}{m}}} dk\ \hbar\omega(k) \left(\frac{1}{e^{\beta\hbar\omega(k)} - 1} + \frac{1}{2} \right) \tag{16.3.8}$$

16.4 1D Diatomic harmonic chain

Often there will be more than one type of atom in a solid, and one shortcoming of the monoatomic chain is that all atoms are treated as identical (not indistinguishable). To fix this, let's consider a chain with the same mass m, but alternating spring constants κ_1 and κ_2 as shown below.

We therefore have two different types of atoms: one type, whose position will be denoted by x_n, has a spring constant κ_1 to the left and κ_2 to the right. The other type, whose position will be denoted by y_n, has a spring constant κ_2 to the left and κ_1 to the right. Note that the unit cell now contains two atoms, so the real space lattice has periodicity a spanning two atoms. The implication of this is that the number of modes is no longer equal to $2N$, the number of masses, but rather N, the number of unit cells.

$$\text{number of modes} = \text{number of unit cells} \tag{16.4.1}$$

The equations of motion now read

$$m\ddot{x}_n = \kappa_2(y_n - x_n) + \kappa_1(y_{n-1} - x_n) \tag{16.4.2}$$
$$m\ddot{y}_n = \kappa_1(x_{n+1} - y_n) + \kappa_2(x_n - y_n) \tag{16.4.3}$$

Once again we propose an mode ansatz

$$x_n = A_x e^{i\omega t - ikna} \tag{16.4.4}$$
$$y_n = A_y e^{i\omega t - ikna} \tag{16.4.5}$$

which possesses the same reciprocal space periodicity encountered in the monoatomic chain.

$$-m\omega^2 A_x e^{-ikna} = \kappa_2(A_y e^{-ikna} - A_x e^{-ikna}) + \kappa_1(A_y e^{-ik(n-1)a} - A_x e^{-ikna}) \tag{16.4.6}$$
$$-m\omega^2 A_y e^{-ikna} = \kappa_1(A_x e^{-ik(n+1)a} - A_y e^{-ikna}) + \kappa_2(A_x e^{-ikna} - A_y e^{-ikna}) \tag{16.4.7}$$

which we can write in matrix form as:

$$-m\omega^2 \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} -(\kappa_1+\kappa_2) & \kappa_2 + \kappa_1 e^{ika} \\ \kappa_1 e^{-ika} + \kappa_2 & -(\kappa_1+\kappa_2) \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} \tag{16.4.8}$$

This is just an eigenvalue problem, and we find that

$$(\kappa_1 + \kappa_2 - m\omega^2)^2 - |\kappa_1 e^{ika} + \kappa_2|^2 = 0 \tag{16.4.9}$$
$$\implies \boxed{m\omega^2 = \kappa_1 + \kappa_2 \pm \sqrt{\kappa_1^2 + \kappa_2^2 + 2\kappa_1\kappa_2\cos(ka)}} \tag{16.4.10}$$

As expected, because each unit cell has 2 atoms, each wavenumber k has two associated normal modes. The dispersion relation is plotted below:

We see that one type of mode, known as **acoustic mode**, behaves similarly to the monoatomic chain. However, we now get a second type of mode known as an **optical mode**. More generally, in D-dimensions, if there are n atoms in each unit cell, there we will obtain $D \cdot n$ degrees of freedoms and thus modes, of which d are acoustic (one for each direction) and the remaining are optical.

Let's calculate the speed of sound in the acoustic modes. The compressibility was given by

$$\beta = -\frac{1}{L}\frac{\partial L}{\partial F} = \frac{1}{Ka} \tag{16.4.11}$$

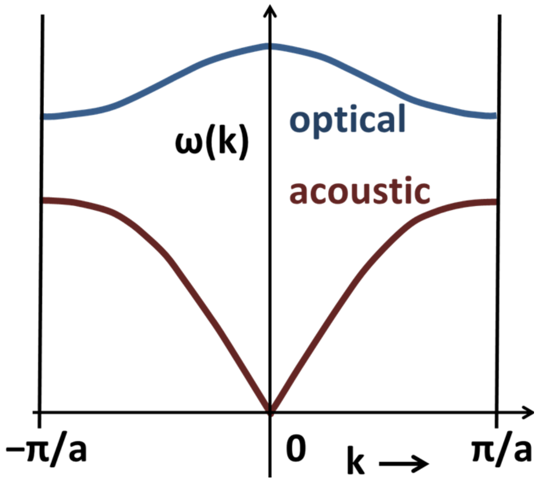


Figure 16.1. Dispersion relation $\omega = \omega(k)$ for a 1D classical diatomic chain

where K is the effective spring constant in the chain, which in the case of two springs in parallel is $K = \frac{\kappa_1\kappa_2}{\kappa_1+\kappa_2}$. Therefore

$$v_s = \sqrt{\frac{1}{2m/a}\frac{\kappa_1+\kappa_2}{a\kappa_1\kappa_2}} = \sqrt{\frac{\kappa_1\kappa_2 a^2}{2m(\kappa_1+\kappa_2)}} \tag{16.4.12}$$

Let's look more closely at the long-wavelength regime. The eigenvalue equation reads

$$-m\omega^2\begin{pmatrix}A_x\\A_y\end{pmatrix} = -(\kappa_1+\kappa_2)\begin{pmatrix}1 & -1\\-1 & 1\end{pmatrix}\begin{pmatrix}A_x\\A_y\end{pmatrix} \tag{16.4.13}$$

which we solve

$$\text{Acoustic mode}: \;\; \omega = 0, \;\begin{pmatrix}A_x\\A_y\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix} \tag{16.4.14}$$

$$\text{Optical mode}: \;\; \omega = \sqrt{\frac{2(\kappa_1+\kappa_2)}{m}}, \;\begin{pmatrix}A_x\\A_y\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix} \tag{16.4.15}$$

The optical mode corresponds to out-of-phase motion, so very high frequency motion (optical), while the acoustic mode corresponds to in-phase motion with lower frequency (acoustic) due to a lower compressibility.

16.5 1D Tight-binding chain

Having looked at classical 1D chains, let us now add quantumness to our discussion. We consider a 1D chain of nuclei with spacing a and position vectors $\mathbf{R}_i$, and electrons modelled by the independent-electron Hamiltonian

$$H = \frac{\mathbf{p}^2}{2m} + \sum_j V(\mathbf{r}-\mathbf{R}_j) \tag{16.5.1}$$

We define $|n\rangle$ to correspond to the state where the electron is localised on the nth nucleus, and assume that they are orthogonal

$$\langle n \,|\, m\rangle = \delta_{nm} \tag{16.5.2}$$

This holds in the atomic limit where the $a \to \infty$, but is pretty bad otherwise. Nevertheless this could still yield good qualitative results so we will proceed and insert the LCAO ansatz

$$|\psi\rangle = \sum_n c_n \,|n\rangle \tag{16.5.3}$$

into the Schroedinger equation

$$\sum_n \langle n \,|\, H \,|\, m\rangle \,\phi_m = E\phi_n \tag{16.5.4}$$

To calculate the matrix elements we split the Hamiltonian

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r} - \mathbf{R}_m) + \sum_{j\neq m} V(\mathbf{r} - \mathbf{R}_j) = H_m + H_{m\neq j} \tag{16.5.5}$$

and find

$$\langle n \,|\, H \,|\, m\rangle = \varepsilon_{\text{atomic}}\delta_{mn} + \sum_{j\neq m} \langle n \,|\, V(\mathbf{r} - \mathbf{R}_j) \,|\, m\rangle \qquad \text{where } \varepsilon_{\text{atomic}} = H_m \,|m\rangle \tag{16.5.6}$$

$\sum_{j\neq m} \langle n \,|\, V(\mathbf{r} - \mathbf{R}_j) \,|\, m\rangle$ will give two types contributions, one when $n = m$ which corresponds to the atom increasing its energy while remaining on site m:

$$V_0 \equiv \sum_{j\neq m} \langle m \,|\, V(\mathbf{r} - \mathbf{R}_j) \,|\, m\rangle \tag{16.5.7}$$

and the other when $n \neq m$ which corresponds to the atom hopping to site n

$$-\,t \equiv \sum_{j\neq m} \langle n \,|\, V(\mathbf{r} - \mathbf{R}_j) \,|\, m\rangle \tag{16.5.8}$$

It should be hard for an electron to hop very far so we can restrict $|n - m| \leq 1$. Finally we have that

$$\sum_{j\neq m} \langle n \,|\, V(\mathbf{r} - \mathbf{R}_j) \,|\, m\rangle \,, \epsilon_{\text{atomic}} = \begin{cases} V_0 & \text{if } n = m \\ -t & \text{if } |n - m| \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{16.5.9}$$

We define the on-site energy to be $\epsilon_0 = \epsilon_{\text{atomic}} + V_0$ and write the Hamiltonian as

$$\boxed{H_{ij} = \epsilon_0 \delta_{i,j} - t(\delta_{i,j+1} + \delta_{i,j-1})} \iff H = \begin{pmatrix} \varepsilon_0 & -t & 0 & 0 & \dots & -t \\ -t & \varepsilon_0 & -t & 0 & \dots & 0 \\ 0 & -t & \varepsilon_0 & -t & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -t & \varepsilon_0 & -t \\ -t & 0 & \dots & 0 & -t & \varepsilon_0 \end{pmatrix} \tag{16.5.10}$$

We make the same wave ansatz as in the classic monoatomic chain

$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_n e^{-ikna} |n\rangle \tag{16.5.11}$$

and find

$$\varepsilon_0 e^{-ikna} - t(e^{-ikn(a-1)} + e^{-ikn(a+1)}) = Ee^{-ikna} \tag{16.5.12}$$

$$\implies \boxed{E(k) = \varepsilon - 2t\cos(ka)} \tag{16.5.13}$$

Newton's equations are second order in time derivatives, so this will provide two frequency solutions. Schroedinger's equation instead is linear in time derivatives so we only obtain one energy solution. This is why we only obtained one energy band with our tight-binding model, rather than two like in the classical chain. For low momenta we can taylor expand (16.5.13)

$$E(k) \approx E_0 + \frac{k^2a^2}{2} \tag{16.5.14}$$

At low frequencies/energies we expect long-wavelength solutions, so free electrons will follow this dispersion relation. We can define an **effective mass**

$$m_{\text{eff}} = \frac{\hbar^2}{2a^2t} \tag{16.5.15}$$

so that the dispersion relation at low frequencies becomes

$$E(k) \sim \frac{\hbar k^2}{2m_{\text{eff}}} \tag{16.5.16}$$

As one should expect, the effective mass is larger as the hopping parameter gets smaller (less hopping implies more inertia). The range of energies, $E_{\text{max}} - E_{\text{min}} = 4t$ is known as the bandwidth of the system.

This solves our problem of the disproportionately large scattering length in the Sommerfeld model. Therefore, the

Let's consider a chain of monovalent atoms where every nucleus gives away one electron from its valence shell. This will produce half-filled bands, with low energy excitations involving the promotion of electrons over the Fermi energy. This means that if one applies the current the Fermi surface can be shifted to the left or the right, thus producing a current. For a chain of diavalent atoms instead, we will obtain a filled band so there will be no low-energy excitations. The heat capacity and the conductivity will therefore be zero.

For a two-orbital model where each unit cell contains two different nuclei then we will obtain two energy bands. If each nucleus gives away three electrons then the lower band will be completely filled, while the upper band will only be half-filled. This means that to investigate transport properties we need only to look at the upper energy band (intraband excitations require massive energies). This explains why when studying materials with several electrons per atom, we do not need to take all of them into account as most non-valence shell electrons will be inert.

A notable exception occurs when two energy bands overlap. Then even a diavalent material will have two partially field bands which will both be active in heat and electrical conduction.

Crystal lattices and reciprocal lattices

17

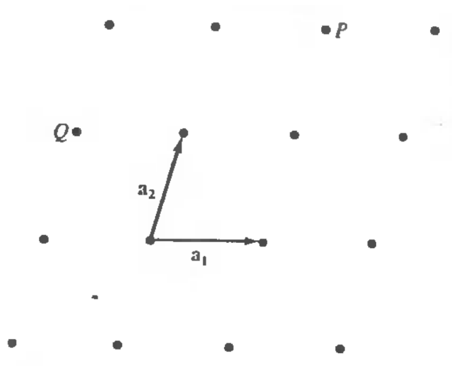
In the previous chapter we saw that the translational symmetry of materials can account for several, such as the propagation of sound modes/phonons, and can provide a better approximation for the specific heat capacity. Nevertheless we were restricted to the 1D case due to its simplicity, and in this chapter we will develop the vocabulary necessary to describe higher dimensional periodic structures. This will be necessary in order to extend our ideas of the 1D chain to the general case of 3D lattices.

17.1 Real space lattices

A **crystal** is defined as an infinite periodic arrangement of its constituent atoms. A crystal's **Bravais lattice**, also known as **real space lattice**/**direct lattice** (DL), is the set of the infinite points defined as integer sums of the crystal **lattice vectors**, which are vectors connecting any two atoms in the crystal. Each point on the real space lattice corresponds thus corresponds to an atom. The **primitive lattice vectors** are the smallest (in magnitude) lattice vectors that span crystal lattice:

$$\text{span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) = \text{DL} \tag{17.1.1}$$

For a Bravais lattice, the primitive lattice vectors always connect two nearest neighbour sites. The number of nearest neighbours to a given lattice point is known as the **coordination number**.



In general, the vector pointing to a point in a Bravais lattice is then given by

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \iff \mathbf{R} = [n_1, n_2, n_3],\ n_i \in \mathbb{Z} \tag{17.1.2}$$

Equivalently, we can define a Bravais lattice as a set of points such that it looks the same viewed from any lattice point. From this definition it is easy to see that the honeycomb lattice

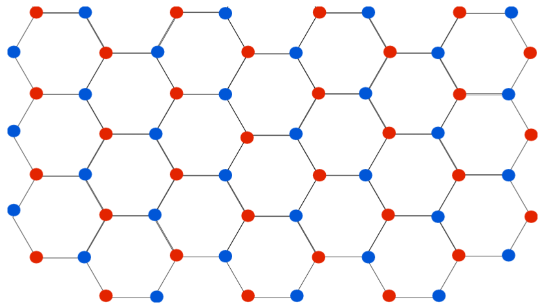
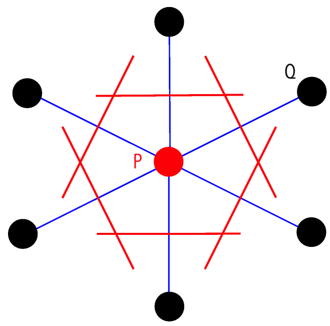


Figure 17.1. The honeycomb lattice is not a Bravais lattice.

is *not* a Bravais lattice, since a red point's environment is not equivalent to a blue point's environment (it is rotated by π). It is however a lattice as it is an infinite periodic structure.

By definition, a periodic structure has a repeating motif, so given a lattice any such repeating cell is known as a **unit cell**. It follows from this definition one should be able tessellate $\mathbb{R}^n$ using the unit cells to obtain the corresponding lattice. Note that unit cells are not unique, this can be seen by choosing a specific unit cell, tessellating the lattice with it, and then translating the lattice by some small distance thus obtaining a new unit cell. Furthermore, a unit cell which contains only one lattice point is known as a **primitive unit cell**. Finally, if the unit cell is made of orthogonal axes then it is a **conventional unit cell** (this does not have to be primitive, but it could be). To top off this definition extravaganza, we define the **Wigner-Seitz cell** (WS cell) as the region around a lattice point that is closer to that lattice point than any other.

A WS cell can be constructed using the Wigner-Seitz construction:



(i) choose a lattice point P around which we will construct the WS cell

(ii) choose a neighbouring lattice point Q and draw the line PQ

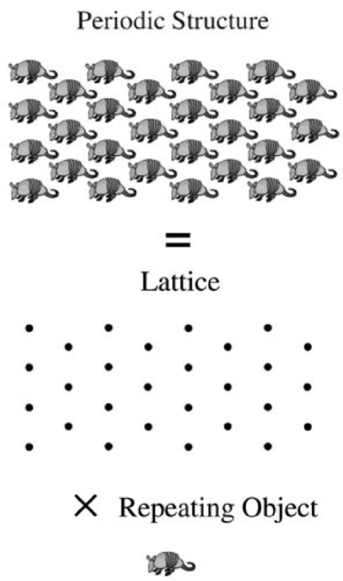
(iii) draw the perpendicular bisector of PQ

(iv) repeat this for all other neighbouring lattice points

(v) the WS cell is the area enclosed by all these perpendicular bisectors.

The WS cell is a primitive unit cell: for obvious reasons it only contains one point, and by the symmetry of its construction it tessellates the lattice. The **Voronoi cell** is a Wigner-Seitz cell for a non-lattice set of points (and it's really important for pizza-delivery).

Since any periodic structure has a repeating motif, it can be viewed as a lattice where each lattice point is occupied by this motif. For example, the following tiling of armadillos can be viewed as a triangular lattice where each site has an armadillo.



The description of objects in a unit cell with respect to the reference lattice point of the unit cell is known as a **basis**. The lattice plus the basis thus defines any periodic structure. By convention, if the reference lattice point is split into the corners of the unit cell, it is counted only once and taken to be on one site only. For example, the honeycomb lattice can be viewed as a triangular lattice with a repeating WS cell enclosed in dashed lines. Each WS cell contains only one lattice point, and has one dark gray and one light gray point inside of it, connected by lines to form a motif. The primitive lattice vectors of the triangular lattice are

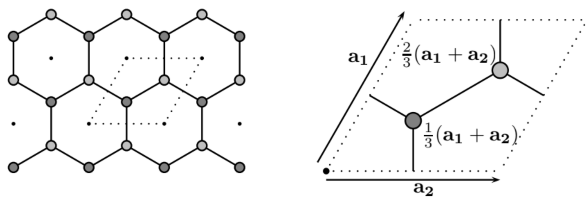
$$\mathbf{a}_1 = a\mathbf{x},\ \mathbf{a}_2 = \frac{a}{2}(\mathbf{x} + \sqrt{3}\mathbf{y}) \tag{17.1.3}$$

so this implies that the basis for the honeycomb lattice is

$$\mathbf{R}_{\text{light}} = \frac{2}{3}(\mathbf{a}_1 + \mathbf{a}_2) \tag{17.1.4}$$

$$\mathbf{R}_{\text{dark}} = \frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2) \tag{17.1.5}$$

Since the reference point of each unit cell has coordinates $[n_1, n_2]$, the basis completely specifies the positions of all points in the hoenycomb lattice.



17.2 3D lattices

In 3D we have the following important unit cells

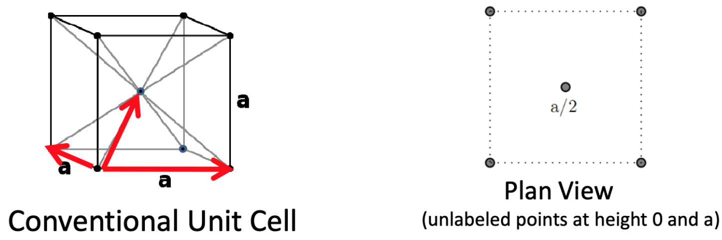
(i) simple cubic unit cell where $a = b = c$

(ii) simple tetragonal unit cell where $a = b \neq c$

(iii) simple orthorhombic unit cell where a, b, c all different

There are two notable 3D lattices with cubic unit cells that do not fall under this category, but are nevertheless important to remember.

The **body-centered cubic lattice (bcc)** is a simple cubic unit cell with the addition of an extra lattice point at the center of the cube.



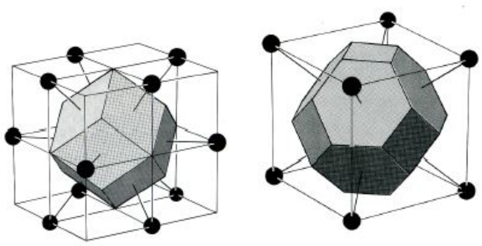
The primitive lattice vectors are

$$\mathbf{a}_1 = [1, 0, 0], \ \mathbf{a}_2 = [0, 1, 0], \ \mathbf{a}_3 = [1/2, 1/2, 1/2] \tag{17.2.1}$$

The points of the bcc lattice are given by the lattice vectors

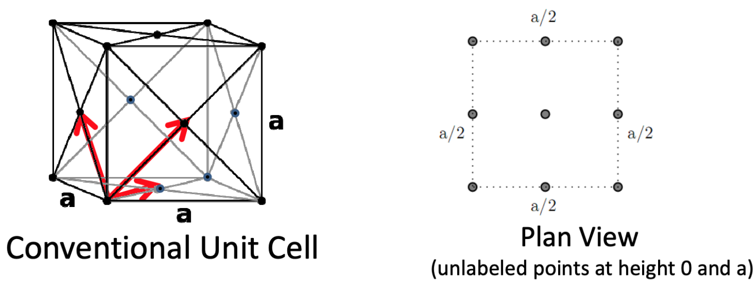
$$\mathbf{R}_{\text{corner}} = [n_1, n_2, n_3], \tag{17.2.2}$$
$$\mathbf{R}_{\text{center}} = [n_1, n_2, n_3] + [1/2, 1/2, 1/2] \tag{17.2.3}$$

from which it follows that the bcc lattice can be viewed as a simple cubic lattice with basis $\mathbf{R}_{\text{center}}$. Also, note that the conventional unit cell contains 2 points, one point in total from the eight corners and another point from the center of the cube. Furthermore, the bcc lattice has a coordination number of 8, 4 nearest neighbours from the plane above, and another 4 from the plane below.



The reason we use the conventional unit cell for the bcc rather than the Wigner-Seitz cell (primitive unit cell) is because the latter, known as a truncated octahedron, is terribly complex.

The **face-centered cubic lattice (fcc)** is a simple cubic unit cell with the addition of an extra lattice point at the center of each of the cube's faces. The primitive lattice vectors of



the fcc lattice are The primitive lattice vectors are

$$\mathbf{a}_1 = [1/2, 1/2, 0] \tag{17.2.4}$$
$$\mathbf{a}_2 = [1/2, 0, 1/2] \tag{17.2.5}$$
$$\mathbf{a}_3 = [0, 1/2, 1/2] \tag{17.2.6}$$

The points of the fcc lattice are given by the lattice vectors

$$\mathbf{R}_{\text{corner}} = [n_1, n_2, n_3], \tag{17.2.7}$$
$$\mathbf{R}_{\text{center}-xy} = [n_1, n_2, n_3] + [1/2, 1/2, 0] \tag{17.2.8}$$
$$\mathbf{R}_{\text{center}-xz} = [n_1, n_2, n_3] + [1/2, 0, 1/2], \tag{17.2.9}$$
$$\mathbf{R}_{\text{center}-yz} = [n_1, n_2, n_3] + [0, 1/2, 1/2] \tag{17.2.10}$$

The fcc lattice has a coordination number of 12. Indeed, the points closest to $[0, 0, 0]$ are given by any point $[\pm 1/2, \pm 1/2, 0]$ and permutations thereof, which are 12 in total.

An important application of the fcc lattice is that if we place a sphere at each site then we obtain the most efficient spherical packing. This was conjectured by Kepler and only formally proven recently by the Flyspeck team in 2014.

17.3 The reciprocal lattice

We define the **reciprocal lattice** to a direct lattice as the set of points with lattice vectors **G** such that

$$e^{i\mathbf{G}\cdot\mathbf{R}_n} = 1 \tag{17.3.1}$$

where $\mathbf{R}_n$ is any direct lattice vectors:

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_3 + n_3\mathbf{a}_3 \tag{17.3.2}$$

To see why (17.3.1) defines a lattice, we suppose that the reciprocal lattice has primitive lattice vectors $\mathbf{b}_i$ satisfying

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij} \tag{17.3.3}$$

which is satisfied whenever

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \tag{17.3.4a}$$
$$\mathbf{b}_2 = 2\pi\frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \tag{17.3.4b}$$
$$\mathbf{b}_3 = 2\pi\frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \tag{17.3.4c}$$

Taking as a definition of the reciprocal lattice primitive vectors, then the reciprocal lattice is indeed a lattice if we can show that any **G** satisfying the definition $e^{i\mathbf{G}\cdot\mathbf{R}_n} = 1$ can be written as

$$\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3, \; m_i \in \mathbb{Z} \tag{17.3.5}$$

Consider any reciprocal space vector **G** with $m_i \in \mathbb{R}$, then we see that

$$1 = e^{i\mathbf{G}\cdot\mathbf{R}} = e^{2\pi(m_1 n_1 + m_2 n_2 + m_3 n_3)} \tag{17.3.6}$$

Since $n_i \in \mathbb{Z}$, it follows that this is only satisfied when $m_i \in \mathbb{Z}$.

Using the expression for the primitive reciprocal lattice vectors we see that the volume of the primitive reciprocal unit cell is

$$\mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = \left(\frac{2\pi}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)}\right)^3 (\mathbf{a}_2 \times \mathbf{a}_3) \cdot \mathbf{a}_1[\mathbf{a}_3 \cdot (\mathbf{a}_1 \times \mathbf{a}_2)] = \frac{(2\pi)^3}{v} \tag{17.3.7}$$

where $v = V/N = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)$ is the volume of the unit cell in direct space.

Interestingly, the reciprocal lattice of fcc is bcc and vice-versa. Indeed more generally the reciprocal lattice is the Fourier transform of the direct lattice. For example, in 1D, setting $R_n = an$ to be the lattice vectors then we may write the mass distribution as a Dirac comb

$$\rho(x) = \sum_n \delta(x - R_n) \tag{17.3.8}$$

Taking its fourier transform we find

$$\mathcal{F}[\rho(x)] = \int dx\, e^{ikx} \rho(x) = \sum_n e^{ikan} = \frac{2\pi}{a} \sum_n \delta(k - G_n) \tag{17.3.9}$$

where $G_n = \frac{2\pi}{a} n$ for $n \in \mathbb{Z}$. We can extend this argument to 3D, taking

$$\rho(\mathbf{r}) = \sum_{\mathbf{R} \in \mathrm{DL}} \delta(\mathbf{r} - \mathbf{R}) \tag{17.3.10}$$

then

$$\mathcal{F}[\rho(\mathbf{r})] = \int d^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}} \rho(\mathbf{r}) = \sum_{\mathbf{R} \in \mathrm{DL}} e^{i\mathbf{k}\cdot\mathbf{R}} = \frac{(2\pi)^3}{v} \sum_{\mathbf{G} \in \mathrm{RL}} \delta(\mathbf{k} - \mathbf{G}_n) \tag{17.3.11}$$

where v is the volume of the primtiive unit cell U. Finally, we can extend our argument to any function with the periodicity of the direct lattice

$$\rho(\mathbf{x}) = \rho(\mathbf{x} + \mathbf{R}), \ \forall \mathbf{R} \in \mathrm{DL} \tag{17.3.12}$$

We find that

$$\mathcal{F}[\rho(\mathbf{r})] = \int d^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}}\rho(\mathbf{r}) = \sum_{\mathbf{R}\in\mathrm{DL}} \int_U d^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}}\rho(\mathbf{r}) \tag{17.3.13}$$

$$= \sum_{\mathbf{R}\in\mathrm{DL}} \int_{U_\mathbf{R}} d^3\mathbf{r}\, e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})}\rho(\mathbf{r}) \tag{17.3.14}$$

$$= \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \int_U d^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}}\rho(\mathbf{r}') \tag{17.3.15}$$

$$= \frac{(2\pi)^3}{v} \sum_{\mathbf{G}\in\mathrm{RL}} \delta(\mathbf{k}-\mathbf{G})S(\mathbf{k}) \tag{17.3.16}$$

where we defined the **structure factor**

$$\boxed{S(\mathbf{k}) = \int_U d^3\mathbf{r}\, e^{i\mathbf{k}\cdot\mathbf{r}}\rho(\mathbf{r})} \tag{17.3.17}$$

and where U is a primitive unit cell with volume v. We have therefore broken down the problem of Fourier transforming a periodic function over all space to that over just one unit cell.

17.4 Fourier analysis on lattices

Here we summarise some other important properties of Fourier analysis on lattices. Let f be a function with the same periodicity of a Bravais lattice: $f(\mathbf{r}+\mathbf{R}) = f(\mathbf{r})$ for all lattice vectors $\mathbf{R}$. It follows that we can expand this function as a superposition of plane waves

$$f(\mathbf{r}) = \sum_{\mathbf{k}\in\mathrm{RL}} \tilde{f}(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{17.4.1}$$

where the Fourier coefficients are given by

$$\tilde{f}(\mathbf{k}) = \frac{1}{v}\int_U e^{i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r})\, d\mathbf{r} \tag{17.4.2}$$

as follows immediately from the orthonormality of $\frac{1}{v}e^{i\mathbf{k}\cdot\mathbf{r}}$ in any unit cell C. Similarly, for a function $\phi(\mathbf{k})$ that is periodic in momentum space then

$$\tilde{\varphi}(\mathbf{k}) = \sum_{\mathbf{r}\in\mathrm{DL}} \varphi(\mathbf{r})e^{i\mathbf{r}\cdot\mathbf{k}},\ \varphi(\mathbf{r}) = \frac{v}{(2\pi)^3}\int e^{-i\mathbf{r}\cdot\mathbf{k}}\tilde{\varphi}(\mathbf{k})\, d\mathbf{k} \tag{17.4.3}$$

We may also need to consider functions that do not have the lattice periodicity, but nevertheless satisfy the Born-von Karman boundary condition

$$f(\mathbf{r}+N_i\mathbf{a}_i) = f(\mathbf{r}),\ \forall \text{ primitive vectors } \mathbf{a}_i, i = 1,2,3 \tag{17.4.4}$$

as often happens with finite crystals. This time the formulae read the same, only that $\mathbf{k}$

goes through the values

$$\mathbf{k} = \sum_{i=1}^{3} \frac{m_i}{N_i} \mathbf{b}_i, \; m_i \in \mathbb{Z} \tag{17.4.5}$$

and $v \to V$ is now the volume of the entire crystal lattice.

We will also often use the following identity for a crystal lattice with N sites and satisfying Born-von Karman conditions:

$$\sum_{\mathbf{R} \in \mathrm{DL}} e^{i\mathbf{k}\cdot\mathbf{R}} = N\delta_{\mathbf{k},0}, \; \mathbf{k} \in \mathrm{FBZ} \tag{17.4.6}$$

To see why, we transform $\mathbf{R} \to \mathbf{R} + \mathbf{R}_0$ where $\mathbf{R}_0$ is another lattice vector. Since we have Born-von Karman conditions the sum overall will not be changed, so we find that

$$\sum_{\mathbf{R} \in \mathrm{DL}} e^{i\mathbf{k}\cdot\mathbf{R}} = e^{i\mathbf{k}\cdot\mathbf{R}_0} \sum_{\mathbf{R} \in \mathrm{DL}} e^{i\mathbf{k}\cdot\mathbf{R}} \tag{17.4.7}$$

The sum therefore vanishes unless $\mathbf{k} \cdot \mathbf{R}_0 = 2\pi n$ for some $n \in \mathbb{Z}$. This must hold for all $\mathbf{R}_0 \in \mathrm{DL}$, implying that $\mathbf{k}$ must be a reciprocal lattice vector. Since it is also restricted to be in the FBZ it follows that $\mathbf{k} = 0$. If $\mathbf{k}$ is allowed to be outside the FBZ then we only require $\mathbf{k}$ to be a reciprocal lattice vector. By similar arguments one can show that

$$\sum_{\mathbf{k} \in \mathrm{FBZ}} e^{i\mathbf{k}\cdot\mathbf{R}} = N\delta_{\mathbf{R},0} \tag{17.4.8}$$

17.5 Lattice planes and miller indices

A **lattice plane** is a plane containing at least three non-collinear lattice points. A **family of lattice planes** is then a set of equidistant parallel lattice planes that contain all the points of a lattice.

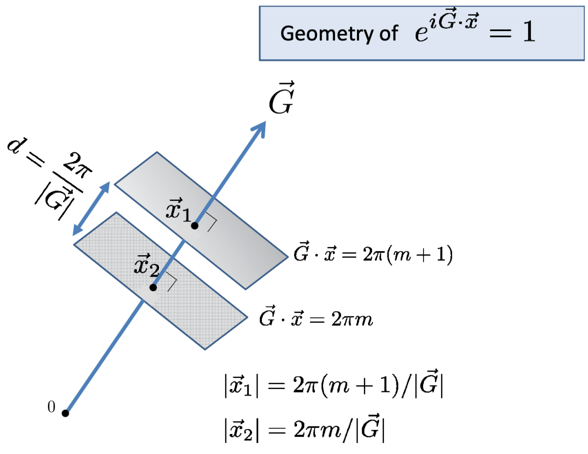
We now claim that the spacing between two neighbouring planes in a family of lattice planes is

$$d = \frac{2\pi}{|\mathbf{G}_{\mathrm{min}}|} \tag{17.5.1}$$

where $\mathbf{G}_{\mathrm{min}}$ is the shortest reciprocal lattice vector normal to these planes. Moreover, the different families of lattice planes are each orthonormal to a different reciprocal lattice vector. Indeed for a given a vector $\mathbf{G}$, $\mathbf{G} \cdot \mathbf{x} = c$ defines a set of points $\mathbf{x}$ forming a plane orthogonal to $\mathbf{G}$. The distance between the plane and the origin can be found by choosing $\mathbf{x}$ to be parallel to $\mathbf{G}$ and taking the modulus of the dot product which yields $|\mathbf{x}||\mathbf{G}|\cos\theta = d|\mathbf{G}| = c$ and thus

$$d = \frac{c}{|\mathbf{G}|} \tag{17.5.2}$$

We can now define a set of planes satisfying $e^{i\mathbf{G}\cdot\mathbf{x}} = 1$ and thus $\mathbf{G} \cdot \mathbf{x} = 2\pi m, \; m \in \mathbb{Z}$, with each m defining a different plane.

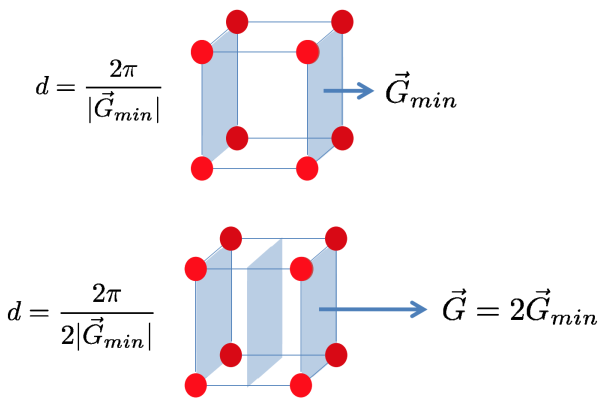


This time the spacing between the planes is

$$d = \frac{2\pi}{|\mathbf{G}|} \tag{17.5.3}$$

To see why, note that letting $\mathbf{x}'$ lie on the m-plane and $\mathbf{x}$ on the $(m+1)$-plane then $\mathbf{G}\cdot(\mathbf{x} - \mathbf{x}') = 2\pi$ and again choosing $\mathbf{x} - \mathbf{x}'$ to lie parallel to $\mathbf{G}$ then we obtain the desired result.

If we take $\mathbf{G}$ to be a reciprocal lattice vector then we see that $\mathbf{R}$ must be direct lattice vectors. Hence the family of planes generated by $e^{i\mathbf{G}\cdot\mathbf{r}} = 1$ will contain some lattice points and will be perpendicular to $\mathbf{G}$ by definition. However, if we want all points to be included, then we need to choose the shortest possible reciprocal lattice vector $\mathbf{G}_{\text{min}}$ parallel to $\mathbf{G}$. Choosing a multiple of $\mathbf{G}_{min}$ would only include a fraction of the lattice points, as shown below



To label families of lattice planes, and their corresponding vector in reciprocal space, we can define reference vectors $\mathbf{b}_i$ which need not be primitive and use **Miller indices** (h, k, l):

$$(h, k, l) = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \tag{17.5.4}$$

These are analogous to the notation we use in real space $[u, v, w]$. It is essential to note that if $\mathbf{a}_i$ are not chosen to be primitive lattice vectors, then the corresponding reference vectors

$\mathbf{b}_i$ vectors will not be reciprocal primitive lattice vectors [1], and consequently not all (h, k, l) will be reciprocal lattice vectors. This is especially important for bcc and fcc lattices where it is often easiest to use the conventional x, y, z axes instead of the reciprocal lattice vectors as basis vectors to simplify calculations. The result is that not all Miller indices produce reciprocal lattice vectors and correspondingly families of lattice planes.

A simple way to evaluate Miller indices of a lattice plane is to compute the intercepts x_1, x_2, x_3 with the axes (defined by the reference vectors) and use the following proportionality

$$h : k : l = \frac{1}{x_1} : \frac{1}{x_2} : \frac{1}{x_3} \tag{17.5.5}$$

These can be easily understood by noting that $\mathbf{G}_{(hkl)} \cdot \mathbf{x} = 2\pi$ defines the cartesian plane $hx_1 + kx_2 + lx_3 = 2\pi$ from which we find that the intercepts $x_1 = \frac{2\pi}{h}$, $x_2 = \frac{2\pi}{k}$ and $x_3 = \frac{2\pi}{l}$.

17.6 Brillouin zone

Any primitive unit cell in reciprocal space is known as a **Brillouin zone** (BZ). The **first Brillouin zone**, in particular, is the region of reciprocal space that is closer to the point at $\mathbf{G} = 0$ than to any other reciprocal lattice point. Analogously, the n**th Brillouin zone** is the region of reciprocal space such that $\mathbf{G} = 0$ is the nth closest reciprocal lattice point. To obtain the nth Brillouin zone one can still use the Wigner-Seitz construction, but this time applied to the second nearest neighbours. Since we can map every point in a given Brillouin zone to the FBZ, the areas of the Brillouin zones must all be the same.

The boundary between two Brillouin zones is known as a Brillouin zone boundary. The distance between two Brillouin zone boundaries is always a reciprocal lattice vector. Also, just like in 1D, the number of momentum states in each Brillouin zone is equal to the number of unit cells.

[1]for example in 1D if we choose the direct lattice vector $R = 2a$ then $G = \frac{2\pi}{R} = \frac{\pi}{a}$ is not the primitive reciprocal lattice vector

X-ray and Neutron scattering

18

18.1 Why scattering?

Scattering is how we see (literally) the world. It is also how we know what DNA, protons, quarks looks like, and in solid state physics it can be used to examine the geometry of crystal lattices. Understanding how scattering experiments can be used to study the will be the goal of this chapter.

18.2 The Laue and Bragg conditions

Choosing the wavelength of the incoming wave to match the size of the scatterer is usually advantageous, and since we are interested in atomic structures which have a size of $\sim$1Å, the corresponding energy scale is

$$E = \frac{hc}{\lambda} \approx 12.3\text{keV} \tag{18.2.1}$$

which is characteristic of X-rays.

Laue condition

Consider an incoming electromagnetic wave with wave-vector $\mathbf{k}$ incident on a scatterer which we can model as a potential $V(\mathbf{r})$. The result of elastic scattering will be a transmitted wave $\mathbf{k}$ and a scattered wave $\mathbf{k}'$. Fermi's golden rule gives the scattering rate

$$\Gamma(\mathbf{k}, \mathbf{k}') = \frac{2\pi}{\hbar} |\langle \mathbf{k} \,|\, V \,|\, \mathbf{k} \rangle|^2 \delta(E_{\mathbf{k}'} - E_{\mathbf{k}}) \tag{18.2.2}$$

where we require $E_{\mathbf{k}'} = E_{\mathbf{k}}$ and $|\mathbf{k}| = |\mathbf{k}'|$ for elastic scattering. The matrix elements are given by

$$\langle \mathbf{k}' \,|\, V \,|\, \mathbf{k} \rangle = \frac{1}{V} \int d^3\mathbf{r}\, e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} V(\mathbf{r}) = \tilde{V}(\mathbf{k} - \mathbf{k}') \tag{18.2.3}$$

where $\tilde{V}(\mathbf{k})$ is the Fourier transform of $V(\mathbf{r})$ and V is the volume of the sample. Assuming the scatterer is a crystal lattice then we can model $V(\mathbf{r})$ as a periodic potential $V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r})$ for all $\mathbf{R} \in$ DL. Consequently using (17.3.11) we find that

$$\tilde{V}(\mathbf{k} - \mathbf{k}') = \frac{(2\pi)^3}{v} \sum_{\mathbf{G} \in \text{RL}} \delta^3(\mathbf{k} - \mathbf{k}' - \mathbf{G}) S(\mathbf{k} - \mathbf{k}') \tag{18.2.4}$$

We can make two important observations from this, firstly that **crystal momentum must be conserved**

Laue condition: there is no scattering unless $\mathbf{k} - \mathbf{k}' \in \mathrm{RL}$, so the **wave-vector transfer** must be a reciprocal lattice vector $\mathbf{G}$.

and secondly the intensity of scattering is proportional to $|S(\mathbf{k} - \mathbf{k}')|^2$.

This Laue condition can be understood alternatively using diffraction. As a starting point we consider two lattice points that are separated by a lattice vector $\mathbf{d}$. An X-ray with wave-vector $\mathbf{k}$ is incident on these two lattice points (since the source is far away we can assume the rays to be parallel) and gets elastically scattered to some wave-vector $\mathbf{k}'$. For the two scattered rays to interfere constructively we require the difference in path lengths they travelled to be an integer multiple of λ.

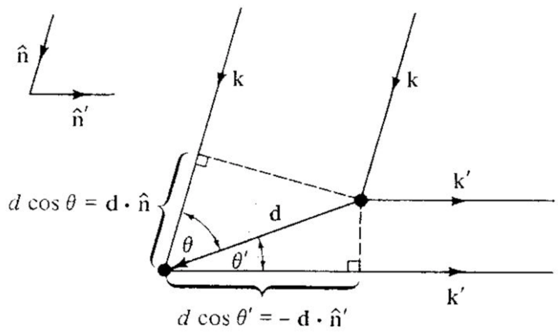


Figure 18.1. Laue condition for constructive interference between two scatterers in a lattice.

From the figure above we see that the incident waves have a path difference of $\mathbf{d} \cdot \hat{\mathbf{k}}$ while the scattered waves have a path difference of $-\mathbf{d} \cdot \hat{\mathbf{k}}'$. Thus the condition for constructive interference is

$$\mathbf{d} \cdot (\hat{\mathbf{k}}' - \hat{\mathbf{k}}) = m\lambda \implies \mathbf{d} \cdot (\mathbf{k}' - \mathbf{k}) = 2\pi m, \; m \in \mathbb{Z} \tag{18.2.5}$$

where we multiplied by $\frac{2\pi}{\lambda}$ and made use of the fact that $|\mathbf{k}| = |\mathbf{k}'|$. Now consider the entire lattice, the condition for constructive interference turns into

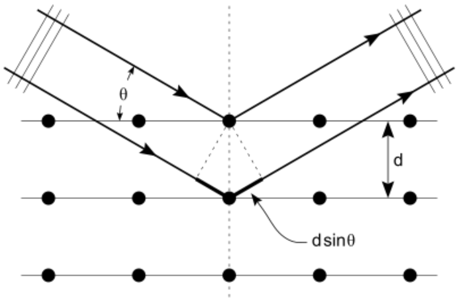
$$\mathbf{R} \cdot (\mathbf{k}' - \mathbf{k}) = 2\pi m, \; m \in \mathbb{Z} \text{ for all lattice vectors } \mathbf{R} \tag{18.2.6}$$

Defining $\mathbf{G} = \mathbf{k}' - \mathbf{k}$ then this is equivalent to $e^{i\mathbf{R}\cdot\mathbf{G}} = 1$ for all $\mathbf{R} \in \mathrm{RL}$ which requires $\mathbf{G}$ to be a reciprocal lattice vector.

Bragg condition

Let us now consider an X-ray undergoing specular scattering (so that the incident and scattered angles are the same) at an angle θ from two successive planes in a given family of lattice planes.

Assuming the planes are separated by d then the condition for constructive interference is



that the path difference, which is $2d\sin\theta$, is an integer multiple of the X-ray wavelength.

$$n\lambda = 2d\sin\theta \tag{18.2.7}$$

This is known as the **Bragg condition**.

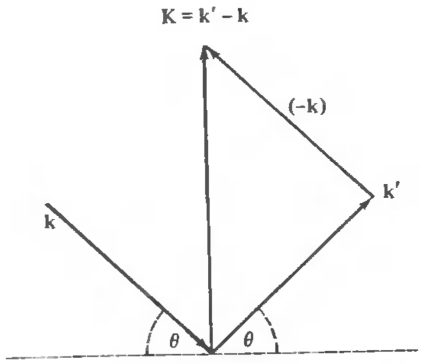
Equivalence of Laue and Bragg conditions

The derivation of the Laue condition has the advantage of not explicitly choosing a specific family of lattice places as well as not assuming that the X-rays will be specularly reflected. Nevertheless it turns out that the Laue and Bragg conditions are equivalent due to the one-to-one correspondence between lattice planes and reciprocal lattice vectors.

To prove this equivalence suppose that the wave-vectors $\mathbf{k}$ and $\mathbf{k}'$ satisfy the Laue condition. Then we find that

$$\mathbf{G}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{k}' - k = -\mathbf{G}\cdot\mathbf{k}' \tag{18.2.8}$$

implying that $\mathbf{k}$ and $\mathbf{G}$ make the same angle with $\mathbf{G}$. If we look at the scattering process



so that the wave-vectors $\mathbf{k},\mathbf{k}'$ lie in the page then we see that Laue scattering with momentum transfer $\mathbf{G}$ can be viewed as Bragg scattering against the family of lattice planes perpendicular to G. Letting this angle be $\pi-\theta$ then

$$\hat{\mathbf{k}}\cdot\hat{\mathbf{G}} = -\sin\theta,\ \hat{\mathbf{k}}'\cdot\hat{\mathbf{G}} = \sin\theta,\ \mathbf{k} = \frac{2\pi}{\lambda}\hat{\mathbf{k}} \tag{18.2.9}$$

The Laue condition with these definitions becomes

$$\mathbf{k}'-\mathbf{k} = \mathbf{G} \implies \frac{2\pi}{\lambda}(\hat{\mathbf{k}}'-\hat{\mathbf{k}}) = \mathbf{G} \tag{18.2.10}$$

and dotting with $\hat{\mathbf{G}}$ then we get that

$$\frac{2\pi}{|\mathbf{G}|} \cdot 2\sin\theta = \lambda \tag{18.2.11}$$

Finally, since $\mathbf{G} = n\mathbf{G}_{\text{min}}$ for some $n \in \mathbb{Z}$ we find that

$$n\lambda = 2d\sin\theta \tag{18.2.12}$$

as desired. Thus the condition for constructive diffraction between lattice planes perpendicular to $\mathbf{G}$ is equivalent to conserving crystal momentum with the momentum transfer being equal to $\mathbf{G}$. In other words, the Bragg scattering from a family of lattice planes perpendicular to $\hat{\mathbf{K}}$ is equivalent to the Laue scattering with momentum transfer $\mathbf{G}$, and the order n of the Bragg peak corresponds to ratio between $|\mathbf{G}|$ and $|\mathbf{G}_{\text{min}}|$.

The Ewald construction

A nice way to visually see if the Laue condition is satisfied (and thus the Bragg condition as well) is via the **Ewald construction**. Consider the reciprocal lattice of a crystal and an arbitrary wave-vector $\mathbf{k}$ as shown below. We draw a circle (or more generally a sphere) of radius k centered at the tip of this wave-vector. Clearly, if there is at least one reciprocal lattice point on the circle/sphere's surface then the Laue condition will be satisfied. Indeed the scattered wave-vector $\mathbf{k}$ lies on the sphere's surface so that $|\mathbf{k}'| = k$ and moreover $\mathbf{k}' - \mathbf{k} = \mathbf{G}$ (since the vector joining any two lattice points is a lattice vector).

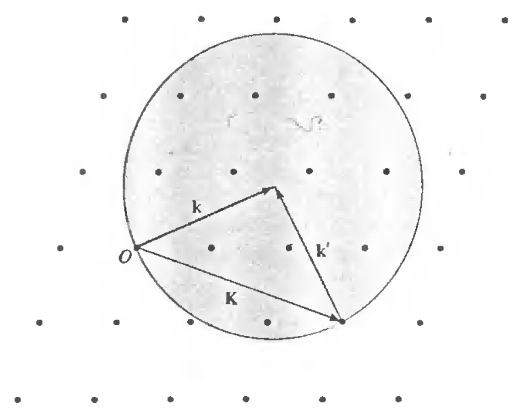


Figure 18.2. Ewald construction for a 2D lattice.

18.3 The scattering amplitude

Depending on what type of wave is being scattered the interaction potential $V(\mathbf{r})$ will take different forms. Generally it is a good approximation to write the scattering potential as a sum of the potential for each individual atom:

$$V(\mathbf{r}) = \sum_{j \in \text{atoms}} V_j(\mathbf{r} - \mathbf{r}_j) \tag{18.3.1}$$

where j runs over all atoms in the system. This ignores the effects that different atoms have on each other. For example, neutrons interact through short range nuclear forces which we approximate as Dirac combs centered at each atom

$$V(\mathbf{r}) \sim \sum_{n\in\text{atoms}} b_n \delta(\mathbf{r} - \mathbf{r}_n) \tag{18.3.2}$$

where b_n is the nuclear scattering length, and $\mathbf{r}_n$ is the position of the nth atom. Assuming b_n is given then

$$\boxed{S(\mathbf{G}) = \sum_{n\in U} b_n e^{i\mathbf{G}\cdot\mathbf{r}_n}} \tag{18.3.3}$$

where n runs over the atoms in the unit cell U.

X-rays, on the other hand, scatter from electrons so instead of interacting through a delta-function peaked at the nuclei, the interaction is mediated through the electron cloud with density $g(\mathbf{r} - \mathbf{r}_n)$. Then the potential experienced by the X-ray wave when encountering atom j is

$$V_j(\mathbf{r} - \mathbf{r}_n) = \sum_{n\in\text{atoms}} Z_n g(\mathbf{r} - \mathbf{r}_n) \tag{18.3.4}$$

and thus the structure factor is

$$\boxed{S(\mathbf{G}) = \sum_{n\in U} e^{i\mathbf{G}\cdot\mathbf{r}} f_n(\mathbf{G})} \tag{18.3.5}$$

where f_n is the **atomic form factor**

$$f_n(\mathbf{G}) = Z_n \int_{\mathbb{R}^3} d^3\mathbf{r}\, e^{i\mathbf{G}\cdot\mathbf{r}} g_n(\mathbf{r}) \tag{18.3.6}$$

In general $f_n(\mathbf{G})$ can be taken as a decreasing function of the deflection angle. Indeed it can be shown that the electron cloud has constant density within a sphere of radius r_0 and vanishes elsewhere then

$$f_j(\mathbf{G}) \sim 3Z_j \left(\frac{\sin(|\mathbf{G}|r_0) - |\mathbf{G}|r_0 \cos(|\mathbf{G}|r_0)}{|\mathbf{G}|^3 r_0^3} \right) \tag{18.3.7}$$

Note that in both neutron and X-ray scattering experiments, the structure factor has the same form: a sum over the atoms in a unit cell of some form factor (constant for neutrons and $\mathbf{G}$-dependent for X-rays) weighed with $e^{i\mathbf{G}\cdot\mathbf{r}}$.

Example 1: simple cubic

Consider for example CsCl crystal which is simple cubic with a basis

$$\text{Cs: } [0,0,0], \qquad \text{Cl: } [1/2, 1/2, 1/2] \tag{18.3.8}$$

Then the structure factor for $\mathbf{G} = (hkl)$ is given by

$$S(\mathbf{G}_{(hkl)}) \equiv S_{(hkl)} = f^{Cs}_{(hkl)} + f^{Cl}_{(hkl)} e^{2\pi i(h+k+l)/2} \tag{18.3.9}$$
$$= f^{Cs}_{(hkl)} + f^{Cl}_{(hkl)} (-1)^{h+k+l} \tag{18.3.10}$$

Example 2: BCC

Pure Cs on the other hand forms a BCC lattice, or alternatively a simple cubic lattice with

$$\text{Cs: } [0, 0, 0], \qquad \text{Cs: } [1/2, 1/2, 1/2] \tag{18.3.11}$$

so that

$$S_{(hkl)} = f^{Cs}_{(hkl)} [1 + (-1)^{h+k+l}] \tag{18.3.12}$$

Note that the structure factor vanishes unless $h + k + l = 2n$ for some $n \in \mathbb{Z}$. This is a **selection rule** on which Miller indices result in scattering. Note that a similar thing can occur in the simple cubic lattice such as in the CsCl crystal if the structure factors of the constituent atoms are very similar. This can happen if the atomic numbers of these atoms are close, since to lowest order $f_{hkl} \sim Z$.

Example 3: FCC

Finally, let's consider a Cu crystal which forms an FCC lattice, which is a simple cubic lattice with basis

$$\text{Cu: } [0, 0, 0], [1/2, 1/2, 0], [1/2, 0, 1/2], [0, 1/2, 1/2] \tag{18.3.13}$$

Therefore

$$S_{(hkl)} = f^{Cu}_{(hkl)} [1 + (-1)^{h+k} + (-1)^{h+l} + (-1)^{k+l}] \tag{18.3.14}$$

from which we find that h, k, l must be either all even or all odd for non-zero scattering.

To understand where the selection rules come from, recall that not all Miller indices for FCC and BCC lattices represent a set of lattice planes. It turns out that only the wave-vectors with Miller indices that correspond to lattice planes can give non-zero structure factors. Indeed we note that for any atom n in a unit cell then $\mathbf{r}_n = \mathbf{R}_n + \mathbf{X}_n$, where $\mathbf{R}_n$ is the vector to the reference lattice point in the cell and $\mathbf{X}_n$ is the position of the atom in the basis. Therefore

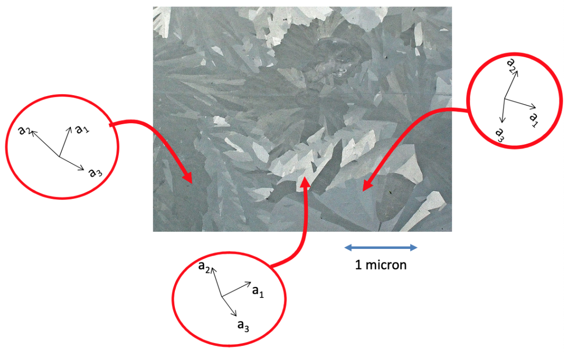
$$S(\mathbf{G}) = \sum_{\mathbf{R} \in U} \sum_{\mathbf{X} \in \text{basis}} f_{\mathbf{X}} e^{i\mathbf{G}\cdot(\mathbf{R}+\mathbf{X})} = \left(\sum_{\mathbf{R} \in U} e^{i\mathbf{G}\cdot\mathbf{R}} \right) \left(\sum_{\mathbf{X} \in \text{basis}} f_{\mathbf{X}} e^{i\mathbf{G}\cdot\mathbf{X}} \right) \tag{18.3.15}$$

which we can write as

$$S = S_{\text{lattice}} \times S_{\text{basis}} \tag{18.3.16}$$

We see that the structure factor vanishes when either the lattice selection rules or the basis selection rules are not satisfied. The lattice selection rules just give the Laue condition for scaterring, while

18.4 Debye-Scherrer powder diffraction



We now consider how one can experimentally determine the crystal structure of a given material. We set up our wave-source which will scatter X-rays against a sample, and set up a detector at a specific solid angle which will measure the light-intensity incident upon it. If one uses a single crystal it is unlikely that for a random orientation of the set-up the crystal will scatter the X-ray towards the detector. Thus there are two options as to how the experiment can be done: one could either rotate the sample with a fixed wave-source or fix the sample and vary the wavelength. Both of these methods are very difficult to achieve and they require virtually perfect crystals. In reality most materials which we consider crystalline are actually poly-crystalline: they contain several domains of crystals oriented in different directions. Although this makes them less pretty visually, for an experimentalist this is good news because one should expect that at some point in the sample there will be a region where the crystal planes are oriented perfectly to achieve detectable scattering. By rotating the sample about the source axis one effectively obtains the scattering pattern from rotating the crystal sample (in a given orientation) around all possible axes.

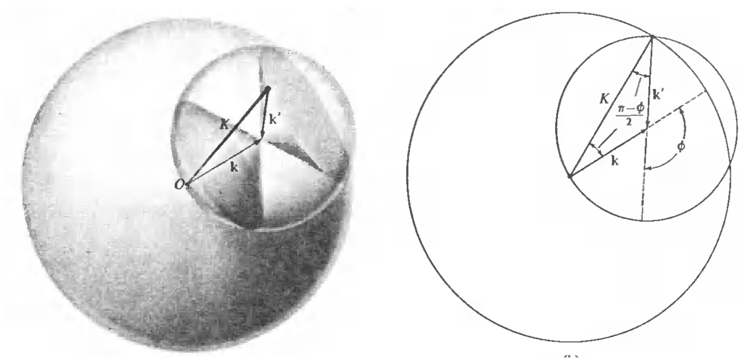


Figure 18.3. The Ewald construction for the Debye-Scherrer powder diffraction. A sphere centered at the origin with radius **K** is formed from the rotation of the polycrystalline sample, and points lying on the interesction between this sphere and the Ewald sphere are suitable scattered wave-vectors $\mathbf{k}'$.

This method is best understood by looking at the Ewald construction. Let's take an incident wave-vector $\mathbf{k}$ and a reciprocal lattice vector $\mathbf{G}$ (which will later become the momentum transfer). We draw the Ewald sphere associated to $\mathbf{k}$ so that any vector joining the tip of $\mathbf{k}$ to a reciprocal lattice point on the Ewald sphere will satisfy the Laue condition. However, since we are effectively rotating the crystal about all axes through some origin O, it could be that some reciprocal lattice point lying outside the Ewald sphere will, at some point in its rotation, be on the Ewald sphere [1]. Therefore we should be looking at the

[1] remember the source is fixed so $\mathbf{k}$ and the Ewald sphere will remain fixed throughout this rotation

sphere with radius **K** centered at O (which gives all the positions of the lattice point at **K** throughout the sample rotation) and its intersection with the Ewald sphere. An intersection always occurs as long as $K < 2k$.

Points lying on this intersection (which in most cases is given by two circles) will at some instant lie on the Ewald sphere and thus be suitable scattered wave-vectors. Consequently any wave-vector $\mathbf{k}'$ joining the tip of $\mathbf{k}$ and any point on the two circular intersections will be a wave-vector satisfying the Laue condition. A **Debye-Scherrer cone** of scattered X-ray waves is obtained, with any wave-vector joining the tip of the $\mathbf{k}$ with any point on this cone's circles representing a possible scattering direction. We will then get different diffraction peaks at different deflection angles 2θ, with $2\theta = 0$ indicating no scattering and $2\theta = \pi$ indicating perfect back-scattering.

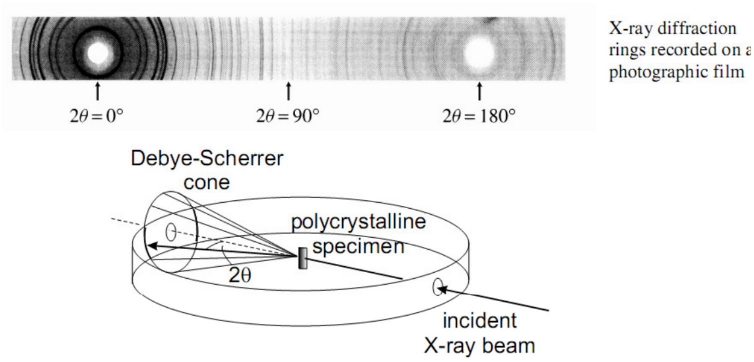


Figure 18.4. The Debye-Scherrer scattering experiment set-up

This method is known as the **Debye-Sherrer powder diffraction method**:

(i) Determine the wavelength of the X-ray

(ii) Obtain the scattering peaks as a function of 2θ, as outlined above.

(iii) Obtain the lattice plane spacing for each peak using $2d \sin\theta = \lambda$.

(iv) Assuming the crystal lattice is cubic, the lattice constant (conventional cubic unit cell's length) a is given by

$$\frac{a^2}{d^2} = h^2 + k^2 + l^2 \equiv N \tag{18.4.1}$$

Since a must be fixed, we are interested in integer sequences of $\frac{a^2}{d^2}$ corresponding to integer values of N. We can find these by looking at the normalised ratios $\frac{d_1^2}{d_n^2}$ where d_m is the lattice spacing obtained from the mth peak. It is clear that

$$\frac{d_1^2}{d_1^1} = 1, \; \frac{d_1^2}{d_2^2} = \frac{N_2}{N_1}, \; \frac{d_1^2}{d_2^2} = \frac{N_3}{N_1}, \; \dots \tag{18.4.2}$$

so by multiplying the sequence of $\left\{\frac{d_1^2}{d_n^2}\right\}$ by an integer (N_1) then one obtains the sequence of values of N.

(vi) Look for selection rules of different values of N and determine the lattice constant $a = dN$.

In ?? we construct a table of selection rules that is often useful. The meaning of each column will become apparent in a second (recall that $\{hkl\}$ represents a family of equivalent Miller indices e.g. $\{100\} = \{(100), (010), (001)\}$).

Scattering Selection Rules

P = Primitive (simple) cubic — All hkl
I = BCC — $h+k+l$ = even
F= FCC — h,k,l all even or all odd

$\{hkl\}$	$N=h^2+k^2+l^2$	Multiplicity	P	I	F
100	1	6	*		
110	2	12	*	*	
111	3	8	*		*
200	4	6	*	*	*
210	5	24	*		
211	6	24	*	*	
---	7	--			
220	8	12	*	*	*
221 , 300	9	24+6	*		
310	10	24	*	*	
311	11	24	*		*
222	12	8	*	*	*
320	13	24	*		
321	14	48	*	*	
---	15	--			
400	16	6	*	*	*

Sequence of N values

P: 1,2,3,4,5,6,8,9, ….. (= all integers excluding 7, 15, 23,…)
I : 2,4,6,8,10,12,14 … (= even integers excluding 28, 60…)
F: 3,4,8,11,12,16,19,20 ….

We see that SC can have any N value that can be written as the sum of three squares (importantly 7, 15 and 23 are not included). BCC on the other hand only has even values of N (not all), and FCC has more difficult patterns.

Example 1: aluminium lattice

Let's consider the aluminium lattice which has the following scattering peaks: We label

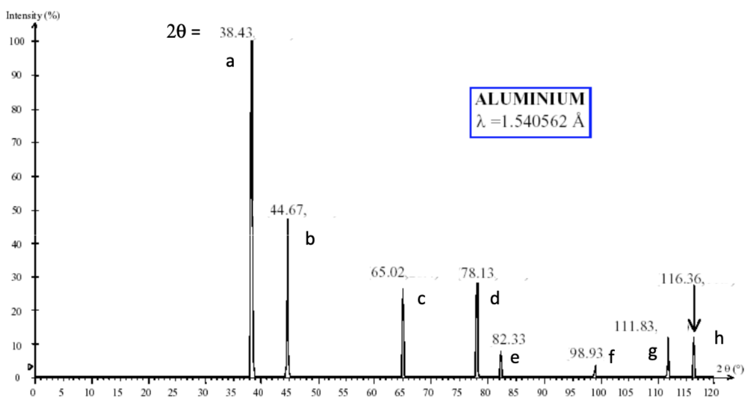


Figure 18.5. X-ray scattering amplitude of aluminium as a function of the deflection angle 2θ.

the peaks as a,b,c,...,h and their corresponding deflection angle 2θ. From these and the X-ray source wavelength λ one can calculate $d = \lambda/(2\sin\theta)$ for each peak. We then set d_a to be the lattice spacing calculated from the first peak and compute d_a^2/d^2. The pattern isn't quite obvious yet but multiplying by three shows that only certain integer values are allowed, which of course correspond to N.

2θ (deg)	d (nm)	d_0^2/d_i^2	N	a (nm)	$\{hkl\}$	BCC	FCC
38.42	0.234	1.000	3.000	0.405	{111}	2	3
44.67	0.203	1.334	4.002	0.405	{200}	4	4
65.02	0.143	2.668	8.004	0.405	{220}	6	8
78.13	0.122	3.668	11.005	0.405	{311}	8	11
82.33	0.117	4.002	12.006	0.405	{222}	10	12
98.93	0.101	5.335	16.006	0.405	{400}	12	16
111.83	0.093	6.336	19.007	0.405	{331}	14	19
116.36	0.091	6.669	20.007	0.405	{420}	16	20

We see that $N = 3, 4, 8, 11...$ produce the peaks, which corresponds to the sequence for an FCC lattice with h, k, l alll even or all odd! Knowing that the lattice is FCC one can then calculate the conventional unit cell size as $a = dN \approx (4.0540 \pm 0.0002)$Å. In some cases the intensity of scattering $I_{(hkl)}$ may also be a useful probe. There are three main factors that determine $|S_{(hkl)}|^2$:

(i) the structure factor $S_{(hkl)}$ determines the phase difference and thus the interference between all the atoms in a primitive unit cell. Indeed if $e^{i\mathbf{G}\cdot(\mathbf{d}-\mathbf{r}_1-\mathbf{r}_2)}$ gives the difference in amplitude between the waves scattered at $\mathbf{r}_1$ and $\mathbf{r}_2$ then the amplitudes of the waves scattered at $\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n$ will be in proportions $e^{i\mathbf{G}\cdot\mathbf{r}_1}, e^{i\mathbf{G}\cdot\mathbf{r}_2}, ..., e^{i\mathbf{G}\cdot\mathbf{r}_n}$. The ray scattered from the entire primitive cell will then have an amplitude given by $S_{(hkl)}$, and it follows that $I_{(hkl)} \propto |S_{(hkl)}|^2$.

(ii) the multiplicity $M_{(hkl)}$ of $\{hkl\}$ determines how many distinct Miller indices give the same scattering angle. It follows that $I_{(hkl)} \propto M_{(hkl)}$.

(iii) the geometric Lorentz factor (dependent on experimental set-up) which can be taken to be constant for intermediate values of θ where most Bragg peaks occur.

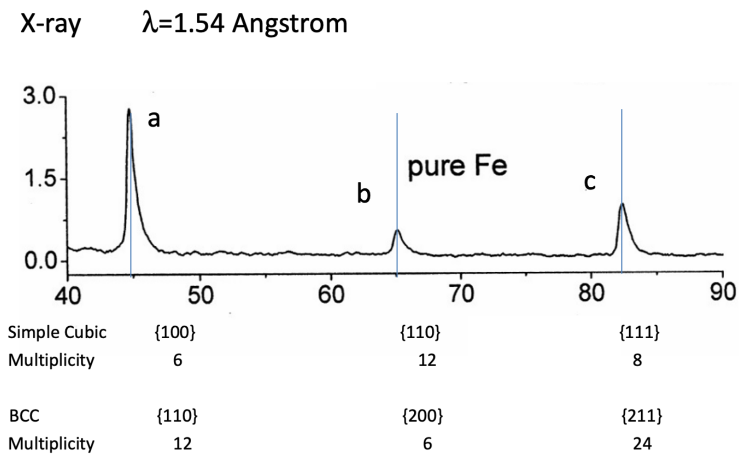
To summarise, we have found that

$$I_{(hkl)} \propto M_{(hkl)} \cdot |S_{(hkl)}|^2 \tag{18.4.3}$$

Example 2: iron

For example, consider the following (bad) scattering data for pure Fe:

This data is quite bad because the peaks are fairly broad and there are not enough of them for an complete comparison between lattice types to be made. Indeed we can go through the typical procedure and find that the allowed values of M are $1, 2, 3$ in the simple cubic case or $2, 4, 6$ in the BCC case. So how do we tell them apart? From the plot we see that the peaks can be ranked in order of height as a, c and b. The form factor is a decreasing function of the deflection angle so the only way c can be taller than b is if $M^c > M^b$. In the

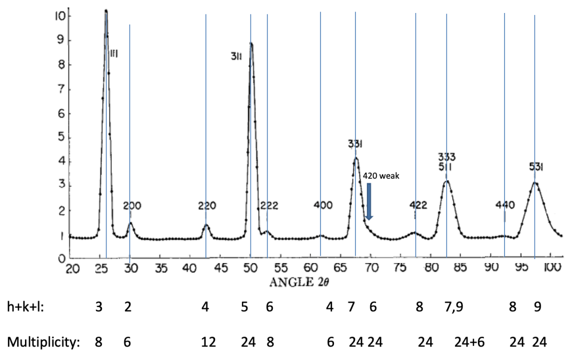


2θ (deg)	N	a (nm)	$\{hkl\}$
44.7	1.000	0.202	100
65.2	2.007	0.202	110
82.70	3.018	0.202	111

simple cubic case $M_{(110)} = 12$ and $M_{(111)} = 8$ which does not fit the intensity patterns. On the other hand, for a BCC lattice $M_{(200)} = 6$ while $M_{(211)} = 24$, as expected.

Example 3: Titanium carbide

As a final example, let's look at a neutron scattering experiment where the form factor is now just a constant. Here is the data for TiC.



2θ (deg)	d (nm)	d_0^2/d_i^2	N	a (nm)	$\{hkl\}$	BCC	FCC
26.0	0.242	1.000	3.000	0.420	111	2	3
30.1	0.210	1.332	3.997	0.420	200	4	4
42.8	0.149	2.631	7.893	0.420	220	6	8
50.2	0.128	3.556	10.668	0.420	311	8	11
52.8	0.123	3.907	11.721	0.420	222	10	12
62.4	0.105	5.303	15.909	0.420	400	12	16
67.6	0.098	6.116	18.347	0.420	331	14	19
70.0	0.095	6.501	19.504	0.420	420	16	20

We can perform the typical analysis and it is easy to see that TiC forms an FCC lattice. Now let's try to understand the basis of TiC, where are the titanium and carbon atoms located? We can define Ti to be at $[0,0,0]$, and let C be at $[u,v,w]$ where u,v,w are unknowns to be computed. The basis structure factor is given by

$$S_{\text{basis}} = \sum_{\text{atoms}\in U} e^{i\mathbf{G}\cdot\mathbf{r}_n} b_n \implies |S_{\text{basis}}|^2 = |b_{\text{Ti}} + b_{\text{C}} e^{2\pi i (hkl)\cdot[uvw]}|^2 \tag{18.4.4}$$

We can make use of the fact that virtually all FCC lattices with a two-atom basis fall under one of the following categories

$$\text{ZnS basis:} \quad [u,v,w] = [1/4,1/4,1/4] \tag{18.4.5}$$
$$\text{NaCl basis:} \quad [u,v,w] = [1/2,1/2,1/2] \tag{18.4.6}$$

With the ZnS basis we find that

$$|S_{\text{basis}}|^2 = |b_{\text{Ti}} + b_{\text{C}}(i)^{h+k+l}|^2 \tag{18.4.7}$$

from which it follows that

$$h+k+l = 4n \implies |S_{\text{basis}}|^2 = |b_{\text{Ti}} + b_{\text{C}}|^2 \tag{18.4.8}$$

$$h+k+l = 4n+2 \implies |S_{\text{basis}}|^2 = |b_{\text{Ti}} - b_{\text{C}}|^2 \tag{18.4.9}$$

$$h+k+l = 2n+1 \implies b_{\text{Ti}}^2 + b_C^2 \tag{18.4.10}$$

The largest peaks therefore occur when $h+k+l$ is even, which certainly does not fit the data where the peaks occur with odd $h+k+l$. Let's check that with the NaCl basis we get the desired behaviour. We find that

$$|S_{\text{basis}}|^2 = |b_{\text{Ti}} + b_{\text{C}}(-1)^{h+k+l}|^2 \tag{18.4.11}$$

from which it follows that

$$h+k+l = 2n \implies |S_{\text{basis}}|^2 = |b_{\text{Ti}} + b_{\text{C}}|^2 \tag{18.4.12}$$

$$h+k+l = 2n+1 \implies |S_{\text{basis}}|^2 = |b_{\text{Ti}} - b_{\text{C}}|^2 \tag{18.4.13}$$

Since the main peaks occur at odd, we find that b_{Ti} and b_{C} must have opposite signs. In conclusion, TiC forms an FCC lattice with basis

$$\text{Ti} = [0,0,0],\ \text{C} = [1/2, 1/2, 1/2] \tag{18.4.14}$$

Electrons in periodic potentials

19

19.1 Bloch's theorem

Consider the following Hamiltonian for electrons moving in a periodic potential $V(\mathbf{r})$

$$H = \frac{\mathbf{p}^2}{2m} + V(\mathbf{r}), \; V(\mathbf{r}+\mathbf{R}) = V(\mathbf{r}) \tag{19.1.1}$$

Here $V(\mathbf{r})$ could represent the electron-ion lattice interaction, with $\mathbf{R}$ being the lattice vector. **Bloch's theorem** states that the eigenstates of H can be written as

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{n,\mathbf{k}}(\mathbf{r}) \;\text{ where }\; \phi_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}) = \phi_{n,\mathbf{k}}(\mathbf{r}) \tag{19.1.2}$$

Proof 1: group theory derivation

This follows immediately from the fact that the Hamiltonian has discrete translational symmetry. Let us define the translation operator

$$T_{\mathbf{R}} = e^{i\mathbf{p}\cdot\mathbf{R}/\hbar} \implies T_{\mathbf{R}}f(\mathbf{r}) = f(\mathbf{r}+\mathbf{R}) \tag{19.1.3}$$

We see that given any reciprocal vectors $\mathbf{R}_i, \mathbf{R}_j$ then

$$[T_{\mathbf{R}_1}, T_{\mathbf{R}_2}] = 0 \tag{19.1.4}$$

and due to the translational invariance of the lattice potential we can also write that

$$[H, T_{\mathbf{R}}] = 0 \tag{19.1.5}$$

Since the translation operator for all lattice vectors and the Hamiltonian commute with each other, we can write down the energy eigenstates as translation operator eigenstates. Since the translation operator is unitary, we can assume that the translation operator eigenvalues are complex phases of the form $e^{i\mathbf{k}\cdot\mathbf{R}}$ so that

$$H\psi_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}}\psi_{n,\mathbf{k}}(\mathbf{r}), \qquad T_{\mathbf{R}}\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n,\mathbf{k}}(\mathbf{r}) \tag{19.1.6}$$

Writing $\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\phi_{n,\mathbf{k}}(\mathbf{r})$ then we must require

$$T_{\mathbf{R}}\psi_{n,\mathbf{k}}(\mathbf{r}) = \psi_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})}\phi_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n,\mathbf{k}}(\mathbf{r}) \tag{19.1.7}$$

This condition also implies that $\phi_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}) = \phi_{n,\mathbf{k}}(\mathbf{r})$, thus proving Bloch's theorem.

There are two important implications of Bloch's theorem. First, the excitations of a periodic potential are fully described in one BZ (usually the FBZ since it is simplest). Secondly, a Hamiltonian with periodic potential conserves crystal momentum, and as such an electron with momentum $\mathbf{k}$ will always remain in that momentum state, it does not get scattered. This solves our mystery of why electrons had such long scattering lengths (hundreds of times larger than the atomic spacing), their wave-functions are well-described by a modulated plane wave which virtually doesn't get scattered, although impurities of course ruin this picture.

Born-Von Karman boundary conditions

Following the first proof it still remains to be shown what the $\mathbf{k}$ vectors are allowed to be. This depends on the boundary conditions of the lattice, the most important being the **Born-Von Karman boundary conditions**. We which imposes periodicity of the wave-function

$$\psi(\mathbf{r}+N_i\mathbf{a}_i) = \psi(\mathbf{r}), \ \forall \mathbf{a}_i \text{ lattice vectors} \tag{19.1.8}$$

where $N = \prod_i N_i$ is the number of unit cells in the crystal. This periodicity of the wave-function can be imposed by taking the crystal lattice to lie in a finite box with periodic boundary conditions. Using Bloch's theorem we see that

$$\psi(\mathbf{r}+N_i\mathbf{a}_i) = e^{iN_i\mathbf{k}\cdot\mathbf{a}_i}\psi(\mathbf{r}) = \psi(\mathbf{r}) \implies \mathbf{k}\cdot\mathbf{a}_i = \frac{2\pi m_i}{N_i}, \ m_i \in \mathbb{Z}, \ \forall i \tag{19.1.9}$$

Letting $\mathbf{b}_i$ be the reciprocal lattice vectors, then let's write

$$\mathbf{k} = \sum_i k_i\mathbf{b}_i \implies k_i = \frac{m_i}{N_i}, \ m_i \in \mathbb{Z}, \ \forall i \tag{19.1.10}$$

Finally, we have found that the allowed $\mathbf{k}$-vectors are given by

$$\mathbf{k} = \sum_i \frac{m_i}{N_i}\mathbf{b}_i \tag{19.1.11}$$

In 3D, the volume of in momentum space afforded to each $\mathbf{k}$ is given by the volume of the parallelepiped whose edges are $\frac{\mathbf{b}_i}{N_i}$

$$\text{vol}(\delta\mathbf{k}) = \frac{\mathbf{b}_1\cdot(\mathbf{b}_2\times\mathbf{b}_3)}{N} \tag{19.1.12}$$

Since the volume of the primitive lattice cell in momentum space is

$$\mathbf{b}_1\cdot(\mathbf{b}_2\times\mathbf{b}_3) = \frac{(2\pi)^3}{V}N \tag{19.1.13}$$

it follows that

$$\text{vol}(\delta\mathbf{k}) = \frac{(2\pi)^3}{V} \tag{19.1.14}$$

This is equal to the volume of the primitive lattice divided by N. Therefore the number of allowed wave-vectors $\mathbf{k}$ is given by the number of total unit cells.

Proof 2: momentum space derivation

Since the potential is periodic it can be expressed as a Fourier series

$$V(\mathbf{r}) = \sum_{\mathbf{q}\in \mathrm{RL}} e^{i\mathbf{q}\cdot\mathbf{r}} V_{\mathbf{q}} \tag{19.1.15}$$

We impose Born-Von Karman conditions so that in the momentum basis the wave-function may be expanded as a superposition of plane waves

$$\psi(\mathbf{r}) = \sum_{\mathbf{q}\in \mathrm{RL}} c_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} \tag{19.1.16}$$

Schroedinger's equation then reads

$$\sum_{\mathbf{q},\mathbf{K}\in \mathrm{RL}} \left(\frac{\hbar^2\mathbf{q}^2}{2m} + V_{\mathbf{K}} e^{i\mathbf{K}\cdot\mathbf{r}} - E\right) c_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} = 0 \tag{19.1.17}$$

$$\implies \sum_{\mathbf{q}'\in \mathrm{RL}} \left[\left(\frac{\hbar^2\mathbf{q}'^2}{2m} - E\right) c_{\mathbf{q}'} + \sum_{\mathbf{K}\in \mathrm{RL}} V_{\mathbf{K}} c_{\mathbf{q}'-\mathbf{K}}\right] e^{i\mathbf{q}'\cdot\mathbf{r}} = 0 \tag{19.1.18}$$

where in the second line we introduced a change of variables $\mathbf{q}' = \mathbf{q} + \mathbf{K}$. Multiplying by $\mathbf{e}^{-i\mathbf{q}\cdot\mathbf{r}}$ and integrating over real space we then obtain that

$$\left(\frac{\hbar^2\mathbf{q}^2}{2m} - E\right) c_{\mathbf{q}} + \sum_{\mathbf{K}\in \mathrm{RL}} V_{\mathbf{K}} c_{\mathbf{q}-\mathbf{K}} = 0 \tag{19.1.19}$$

Note that each of the N wave-vectors $\mathbf{q}$ in the FBZ will produce an equation coupling $c_{\mathbf{q}}$ to its representatives $c_{\mathbf{q}+\mathbf{K}}$, $\forall \mathbf{K} \in$ RL in the other BZs. The end result is that we obtain N distinct equations, and generally there will be more than one solution. Therefore each eigenstate can be labelled by $\mathbf{k} \in$ FBZ and n, the band number identifying which solution $\{c^n_{\mathbf{k}}\}$ is used:

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}\in \mathrm{RL}} c^n_{\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} = e^{i\mathbf{k}\cdot\mathbf{r}} \phi_{n,\mathbf{k}}(\mathbf{r}), \qquad \phi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}\in \mathrm{RL}} c^n_{\mathbf{k}+\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \tag{19.1.20}$$

Finally, note that

$$\phi_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}) = \sum_{\mathbf{G}\in \mathrm{RL}} c^n_{\mathbf{k}+\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} e^{i\mathbf{G}\cdot\mathbf{R}} = \phi_{n,\mathbf{k}}(\mathbf{r}) \tag{19.1.21}$$

as desired.

19.2 The Kronig-Penney model

We begin by looking at some general properties of electrons in periodic 1D potentials. Consider a lattice of ions sitting at the minima of a periodic potential $U = \sum_{n=-\infty}^{\infty} v(x-na)$ which are taken to be zero. We also assume that $v(x)$ is an even function for simplicity. The potential U therefore corresponds to a superposition of single-ion potentials centered at na for $n \in \mathbb{Z}$ which vanish if $|x| \geq (n+1/2)a$. This is known as the **Kronig-Penney model**.

To evaluate the resulting energy spectrum we need to solve the Schrodinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\phi + \sum_{n=-\infty}^{\infty} v(x-na)\phi = E\phi \tag{19.2.1}$$

This is greatly facilitated by the Bloch's theorem, which requires

$$\psi(x+a) = e^{ika}\psi(x) \implies \psi'(x+a) = e^{ika}\psi'(x) \tag{19.2.2}$$

Now consider the single-ion Schrodinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\phi + v(x)\phi = E\phi \tag{19.2.3}$$

This is simply a scattering problem so we can take a plane wave incoming from the left, $\phi_I = e^{ikx}$ in $x < a/2$, which scatters into $v(x)$ producing a reflected wave $\phi_R = re^{-ikx}$ in $x < a/2$ and a transmitted wave $\phi_T = te^{ikx}$ in $x > a/2$. We therefore try the ansatz

$$\phi_L = \begin{cases} e^{iKx} + r^{-iKx}, & x < -a/2 \\ te^{iKx}, & x > a/2 \end{cases} \quad , \; E = \frac{\hbar^2 K^2}{2m} \tag{19.2.4}$$

By inversion symmetry, we can consider the reverse situation of a plane wave incoming from the right and find

$$\phi_R = \begin{cases} te^{iKx}, & x < -a/2 \\ e^{-iKx} + r^{iKx}, & x > a/2 \end{cases} \quad , \; E = \frac{\hbar^2 K^2}{2m} \tag{19.2.5}$$

Note that these are two independent solutions of the Schrodinger equation with energy $E = \frac{\hbar^2 K^2}{2m}$, as can be easily proved by evaluating the Wronskian $W(\phi_L, \phi_R)$. Therefore the general solution to the TISE can be expressed as a linear superposition of ϕ_L and ϕ_R

$$\phi(x) = A\phi_L(x) + B\phi_R(x) \tag{19.2.6}$$

But this is the solution to the single-ion Schrodinger equation, how do we relate it to the general problem of a periodic potential? Note that (19.2.1) and (19.2.3) are the same in the region $|x| \leq a/2$, so by the uniqueness theorem (19.2.6) must be the restriction of $\psi(x)$ to the neighborhood of the ion at the origin:

$$\psi(x) = A\phi_L(x) + B\phi_R(x), \; |x| \leq \frac{a}{2} \tag{19.2.7}$$

We can then use Bloch's theorem to find the wave-function in the other ion's neighborhoods, so that in general

$$\psi(x+na) = e^{ikna}[A\phi_L(x) + B\phi_R(x)], \; |x| \leq \frac{a}{2} \tag{19.2.8}$$

Now it remains to find what values of K are allowed. In general we should expect the presence of band gaps, values of E where there are no solutions to the TISE. To understand why, consider the free electron model. Here we expect to get Without a potential both solutions are degenerate with each other, but as we turn on the potential $V(x) = V(x+a)$, one of these will have its peaks aligned with the maxima of V, while the other will have its maxima aligned where V vanishes. This difference leads to a separation in energy and a gap opening.

In our case, we can impose the Bloch conditions on (19.2.7). We find that $\psi(x+a) = e^{ika}\psi(x)$ requires

$$A[te^{iKa/2} - e^{ika}(e^{-iKa/2} + re^{iKa/2})] = B[te^{ika}e^{iKa/2} - e{-}iKa/2 - re^{iKa/2}] \tag{19.2.9}$$

and $\psi'(x+a) = e^{ika}\psi'(x)$ requires

$$A[te^{iKa} - e^{ika}(e^{-iKa/2} - re^{iKa/2})] = B[-te^{iKa/2}e^{ika} + e^{-iKa/2} - re^{iKa/2}] \tag{19.2.10}$$

Taking the sum of these two we get

$$A(te^{iKa/2} - e^{ika}e^{-iKa/2}) = -2rBe^{iKa/2} \tag{19.2.11}$$

while their difference yields

$$Are^{iKa/2} = -2Be^{-ika}(te^{iKa/2}e^{ika} - e^{-iKa/2}) \tag{19.2.12}$$

Substituting the latter into the first and simplifying we find that

$$\boxed{\cos ka = \frac{t^2 - r^2}{2t}e^{iKa} + \frac{1}{2t}e^{-iKa}} \tag{19.2.13}$$

We are still not done since t and r are not independent. Indeed note that given two solutions to Schrodinger's equation, then

$$W'(\phi_1, \phi_2) = \phi_1''(x)\phi_2(x) - \phi_1(x)\phi_2''(x) = \frac{2m(U-E)}{\hbar^2}\phi_1\phi_2 - \frac{2m(U-E)}{\hbar^2}\phi_2\phi_1 = 0 \tag{19.2.14}$$

The Wronskian for the solutions ϕ_L and ϕ_L^* is

$$W(\phi_L, \phi_L^*) = \begin{cases} 2iK|t|^2, \ x \le -a/2 \\ 2iK(1-|r|^2), \ x \ge a/2 \end{cases} \tag{19.2.15}$$

Since W cannot have any x-dependence we require

$$|r|^2 + |t|^2 = 1 \tag{19.2.16}$$

as expected from probability conservation. Similarly evaluating the Wronskian for ϕ_L and ϕ_R^* we find

$$W(\phi_L, \phi_R^*) = \begin{cases} -2iKt^*r, \ x \le -a/2 \\ 2iKtr^*, \ x \ge a/2 \end{cases} \implies t = |t|e^{i\Delta}, \ r = \pm i|r|e^{i\Delta} \tag{19.2.17}$$

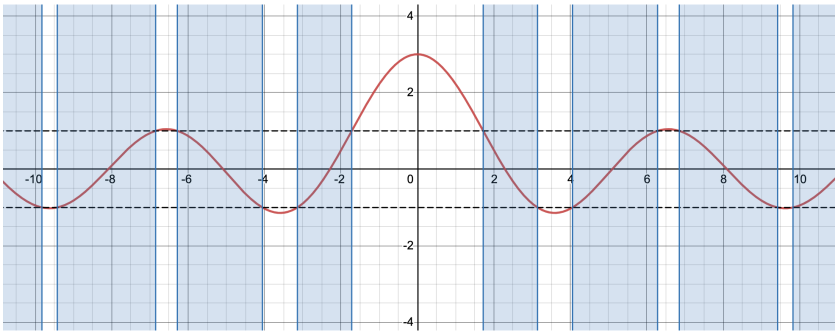
Substituting these conditions into (19.2.13) we find

$$\boxed{\cos ka = \frac{\cos(Ka+\Delta)}{|t|}} \tag{19.2.18}$$

This fully solves the problem, given the form of $v(x)$ one can find the values of $|t|$ and Δ by imposing necessary boundary conditions, and then one can find for which energies $E = \frac{\hbar^2 K}{2m}$ we get a solution and the corresponding crystal momentum k. It is important to note that $-1 < \cos ka < 1$ so that a solution will only exist for those values of K where

$$-|t| < \cos(Ka+\Delta) < |t| \tag{19.2.19}$$

An example is plotted below



For very small values of $|t|$ (very strong potential) there will be very narrow regions of allowed K. On the other hand for very large values of $|t|$ (very weak potential) there will be very narrow regions of forbidden K, which are **band gaps**. These occur near the peaks of $\cos(Ka+\Delta)$ which are larger than $|t|$, but since $|t| \approx 1$ these regions will be narrow. Indeed the width of these band gaps can be found by setting $|r|/|t| \ll 1$, $\Delta \approx 0$, and finding the distance between the solutions to $\cos(Ka) = |t| \implies \cos^2(Ka) = |t|^2$. Thus we taylor expand $\cos^2(Ka)$ about $n\pi$ to get

$$1 - (Ka - n\pi)^2 = |t|^2 \implies \frac{\hbar^2 K^2}{2m} = \frac{\hbar^2}{2ma^2}(\pm|r| + n\pi)^2 \tag{19.2.20}$$

The distance between these two solutions is the band gap centered at $n\pi$

$$\Delta E_{\text{gap}} \approx 2n\pi \frac{\hbar^2}{ma^2}|r| \tag{19.2.21}$$

which is indeed very small.

For a more concrete example, consider the case where $v(x) = v_0\delta(x)$. Here we get a Dirac comb of potentials centered at the ions. Focusing on the region $|x| < a/2$, we obtain a boundary condition by integrating the TISE:

$$\psi'(0^+) - \psi'(0^-) = -\frac{2mv_0}{\hbar^2}\psi_L(0) \tag{19.2.22}$$

We must also require that $\psi(x)$ be continuous so that

$$\psi(0^+) = \psi(0^-) \tag{19.2.23}$$

The Hamiltonian for $x \neq 0$ is just a free-particle Hamiltonian so we can use the plane-wave ansatz

$$\psi(x) = \begin{cases} e^{iKx} + r^{-iKx}, \ x < 0 \\ te^{iKx}, \ x > 0 \end{cases} \quad , \ E = \frac{\hbar^2 K^2}{2m} \tag{19.2.24}$$

We then find that

$$1 + r = t, \ \text{ and } \ iK(1-r) = iKt - \frac{2mv_0}{\hbar^2} t \tag{19.2.25}$$

and after a bit of algebra this yields

$$r = -\frac{1}{1 - i\frac{\hbar^2 K}{mv_0}}, \quad t = \frac{i\frac{\hbar^2 K}{mv_0}}{1 - i\frac{\hbar^2 K}{mv_0}} \tag{19.2.26}$$

Letting $t = |t|e^{i\Delta}$ and $r = \pm i|r|e^{i\Delta}$ then we see that

$$\frac{t}{r} = -i\frac{\hbar^2 K}{mv_0} = \pm i\frac{|t|}{|r|} = \pm i\frac{\hbar^2 K}{mv_0} \tag{19.2.27}$$

so we should take $r = -i|r|e^{i\Delta}$. We then see that

$$t + 1 + r \implies |t|^2 e^{2i\Delta} = 1 - |r|^2 e^{2i\Delta} - 2i|r|e^{i\Delta} \implies |r| = -\sin\Delta, \ |t| = \cos\Delta \tag{19.2.28}$$

So $\cot\Delta = -\frac{\hbar^2 K}{mv_0}$ and

$$\cos\Delta\cos(ka) = \cos(Ka + \Delta) \implies \boxed{\cos(ka) = \cos(Ka) + \frac{mv_0}{\hbar^2 K}\sin(Ka)} \tag{19.2.29}$$

19.3 Nearly-free electron model: electrons in weak periodic potentials

We consider a free-electron system $H_0 = \frac{\mathbf{p}^2}{2m}$ with and treat the ion-lattice potential as a weak periodic perturbation $V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R})$. The eigenstates of H_0 are Bloch states $|\mathbf{k}\rangle$ such that

$$H_0 |\mathbf{k}\rangle = \frac{\hbar^2 k^2}{2m} |\mathbf{k}\rangle = \varepsilon^0_{\mathbf{k}} |\mathbf{k}\rangle \tag{19.3.1}$$

To first order in perturbation theory we find that

$$\varepsilon_{\mathbf{k}} = \varepsilon^0_{\mathbf{k}} + \underbrace{\langle \mathbf{k} \,|\, V \,|\, \mathbf{k} \rangle}_{V_0} \tag{19.3.2}$$

which is just a constant offset V_0. Second order perturbation theory is more interesting

$$\varepsilon_{\mathbf{k}} = \varepsilon^0_{\mathbf{k}} + V_0 + \sum_{\mathbf{k}' \neq \mathbf{k}} \frac{|\langle \mathbf{k}' \,|\, V \,|\, \mathbf{k} \rangle|^2}{\varepsilon^0_{\mathbf{k}} - \varepsilon^0_{\mathbf{k}'}} \tag{19.3.3}$$

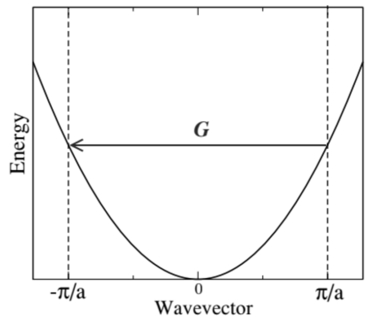
However, Laue's condition requires $\langle \mathbf{k}' \,|\, V \,|\, \mathbf{k} \rangle = \delta_{\mathbf{k}-\mathbf{k}',\mathbf{G}}$. Therefore the second order correction is

$$\varepsilon_{\mathbf{k}}^2 = \sum_{(\mathbf{G}\neq 0)\in \mathrm{RL}} \frac{|\langle \mathbf{k}+\mathbf{G} \,|\, V \,|\, \mathbf{k} \rangle|^2}{\varepsilon_{\mathbf{k}}^0 - \varepsilon_{\mathbf{k}+\mathbf{G}}^0} \tag{19.3.4}$$

This is quite problematic because the denominator blows up due to degeneracies:

$$|\mathbf{k}| = |\mathbf{k}+\mathbf{G}| \tag{19.3.5}$$

These denegeracies occur on the Brillouin zone boundaries, as shown below for the simple 1D case where $G = -\frac{2n\pi}{a}$:



To account for this degeneracy one must use degenerate perturbation theory and diagonalise the Hamiltonian in the degenerate subspace. Letting

$$|\psi\rangle = \phi_{\mathbf{k}} |\mathbf{k}\rangle + \phi_{\mathbf{k}+\mathbf{G}} |\mathbf{k}+\mathbf{G}\rangle \tag{19.3.6}$$

then the Hamiltonian matrix elements in the degenerate subspace are

$$\langle \mathbf{k} \,|\, H \,|\, \mathbf{k} \rangle = \epsilon_{\mathbf{k}}^0 + V_0 \tag{19.3.7}$$
$$\langle \mathbf{k} \,|\, H \,|\, \mathbf{k}+\mathbf{G} \rangle = \epsilon_{\mathbf{k}}^0 + V_{\mathbf{G}}^* \tag{19.3.8}$$
$$\langle \mathbf{k}+\mathbf{G} \,|\, H \,|\, \mathbf{k} \rangle = \epsilon_{\mathbf{k}}^0 + V_{\mathbf{G}} \tag{19.3.9}$$
$$\langle \mathbf{k}+\mathbf{G} \,|\, H \,|\, \mathbf{k}+\mathbf{G} \rangle = \epsilon_{\mathbf{k}+\mathbf{G}}^0 + V_0 \tag{19.3.10}$$

Here we defined the Fourier coefficients

$$V_{\mathbf{q}} = \frac{1}{L^3} \int d\mathbf{r}^3 V(\mathbf{r}) e^{-i\mathbf{q}\cdot\mathbf{r}} \tag{19.3.11}$$

Simple case: k on the BZ boundary

Suppose that $\mathbf{k}$ sits exactly on a BZ boundary. Then in the 1D case we have that $G = -\frac{2n\pi}{a}$ and thus the degenerate space Hamiltonian takes the form

$$H_{\mathrm{deg}} = \begin{pmatrix} \epsilon_{\mathbf{k}}^0 + V_0 & V_{\mathbf{G}}^* \\ V_{\mathbf{G}} & \epsilon_{\mathbf{k}}^0 + V_0 \end{pmatrix} \tag{19.3.12}$$

This can be easily diagonalised to yield the energies to first order in degenerate perturbation theory

$$\epsilon_{\mathbf{k}}^{\pm} = \epsilon_{\mathbf{k}}^{0} + V_0 \pm |V_{\mathbf{G}}| \tag{19.3.13}$$

This perturbation opens a gap $\Delta = 2|V_{\mathbf{G}}|$ near the Brillouin zone boundary, while in the rest of the Brillouin zone the dispersion relation will look roughly parabolic (as predicted by the free-electron model).

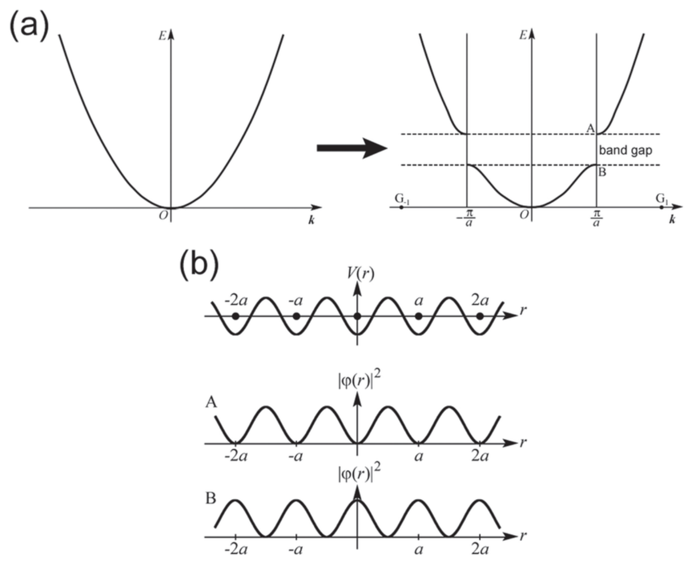
We can understand the opening of this gap more qualitatively by looking at the eigenstates of H_{deg}:

$$\psi_{\mathbf{k}}^{\pm}(x) = \frac{1}{\sqrt{2L}}(e^{ikx} \pm e^{i(k+G)x}) \tag{19.3.14}$$

and so the probability density is

$$|\psi_{\mathbf{k}}^{\pm}(x)|^2 = \frac{1}{L}[1 \pm \cos(Gx)] \tag{19.3.15}$$

We see that the the probability density of the plus state has maxima concentrated near the maxima of $|V_G|\cos(Gx)$, while the minus state. Therefore the electron in the plus state is more likely to be in the high-potential regions of the lattice while the electron in the minus state is likely to be in the low-potential regions. All other sinusoids in the Fourier expansion of V do not contribute as they are out of phase with the probability amplitudes which have wave-number G. The result is that a gap of $2|V_G|$ will open up.



General case: k close the BZ boundary

Suppose that k is a distance δk from the BZ boundary so that $k = \frac{n\pi}{a} + \delta k$ for some $n \in \mathbb{Z}$. The degenerate subspace Hamiltonian now reads

$$H_{\text{deg}} = \begin{pmatrix} \epsilon_{\mathbf{k}}^{0} & V_{\mathbf{G}}^{*} \\ V_{\mathbf{G}} & \epsilon_{\mathbf{k}+\mathbf{G}}^{0} \end{pmatrix}, \; G = -\frac{2n\pi}{a} \tag{19.3.16}$$

where we absorbed V_0 into $\epsilon^0_{\mathbf{k}}$ (this is equivalent to shifting the potential so that its zero-wavenumber mode, the value it oscillates about, is zero). The energy levels are found by solving the secular equation

$$(\epsilon^0_{\mathbf{k}} - \epsilon_{\mathbf{k}})(\epsilon^0_{\mathbf{k}+\mathbf{G}} - \epsilon_{\mathbf{k}}) = |V_{\mathbf{G}}|^2 \tag{19.3.17}$$

We get

$$\epsilon_{\mathbf{k}} = \frac{1}{2}\left[\epsilon^0_{\mathbf{k}} + \epsilon^0_{\mathbf{k}+\mathbf{G}} + 2V_0 \pm \sqrt{(\epsilon^0_{\mathbf{k}} - \epsilon^0_{\mathbf{k}})^2 + 4|V_{\mathbf{G}}|^2}\right] \tag{19.3.18}$$

Substituting

$$\epsilon^0_{\mathbf{k}} = \frac{\hbar^2}{2m}(K+\delta k)^2,\ \epsilon^0_{\mathbf{k}+\mathbf{G}} = \frac{\hbar^2}{2m}(-K+\delta k)^2 \tag{19.3.19}$$

we get that

$$\epsilon_{\mathbf{k}} = \frac{\hbar^2}{2m}(K^2+\delta k^2) \pm |V_{\mathbf{G}}|\sqrt{1+\left(\frac{\hbar^2}{m^2}\frac{K\delta k}{|V_{\mathbf{G}}|}\right)} \tag{19.3.20}$$

$$\approx \frac{\hbar^2}{2m}(K^2+\delta k^2) \pm |V_{\mathbf{G}}|\left(1+\frac{\hbar^2}{2m^2}\frac{K^2\delta k^2}{|V_{\mathbf{G}}|^2}\right) \tag{19.3.21}$$

The Taylor expansion is valid when $K\delta k \ll |V_{\mathbf{G}}|$, so for large values of n we will need k to be closer and closer to the BZ boundary. If this assumption is satisfied then

$$\epsilon^{\pm}_{\mathbf{k}} = \frac{\hbar^2 n^2\pi^2}{2ma^2} \pm |V_{\mathbf{G}}| + \frac{\hbar^2\delta k^2}{2m}\left(1 \pm \frac{\hbar^2 n^2\pi^2}{ma}\frac{1}{|V_{\mathbf{G}}|}\right),\ G = -\frac{2n\pi}{a} \tag{19.3.22}$$

The band structure is shown below Letting ϵ then the electron effective mass is

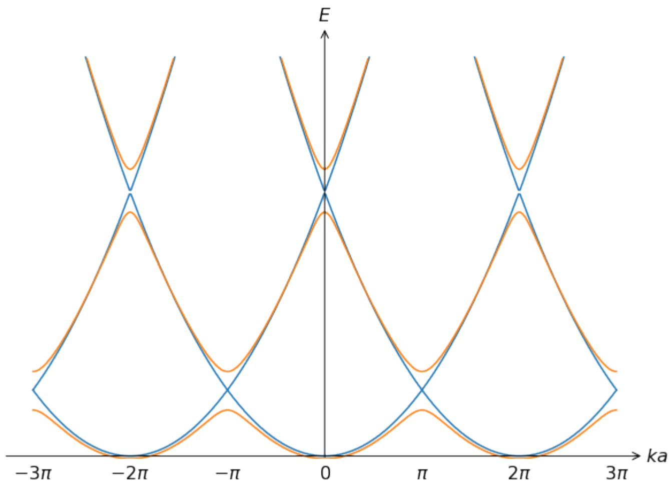


Figure 19.1. The band structure in the nearly free electron model looks parabolic due to small contributions from perturbation theory. On the other hand, near the BZ boundaries a band gap opens due to hybridisation of plane wave forming ungerade and gerade orbitals.

$$m_{\text{eff}} = \frac{m}{\left|1 \pm \frac{\hbar^2 n^2 \pi^2}{ma|V_{\mathbf{G}}|}\right|} \tag{19.3.23}$$

Example: silicon carbide

We show below the Brillouin zone of an FCC lattice, which is equivalent to the Wigner-Seitz cell of a BCC lattice. Then

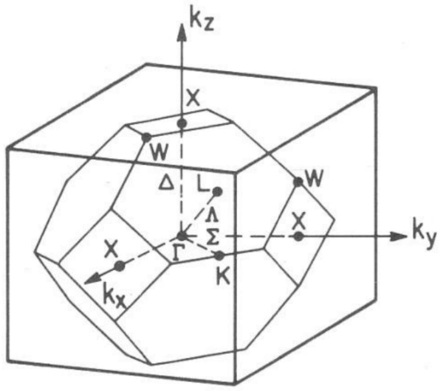


Figure 19.2. Brillouin zone of an FCC lattice

Lets look a the dispersion of pure silicon which forms an FCC lattice with basis (zincblende structure)

$$\text{Si at } [0, 0, 0] \text{ and } \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right] \tag{19.3.24}$$

and silicon carbide which also forms an FCC lattice with basis

$$\text{Si at } [0, 0, 0] \text{ and C at } \left[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right] \tag{19.3.25}$$

which are shown below Interestingly the dispersion relations look very similar for both materials, especially considering the first four bands. Indeed we can see that near the origin Γ the dispersion looks parabolic for both materials. However SiC exhibits a gap opening at X as predicted by our nearly-free electron model, while silicon does not.

The reciprocal lattice vector connecting X on one side of the BZ to X on the other side (that is, -X) is (200). We see that

$$S_{(200)} = S_{\text{FCC}} \cdot f_{\text{Si}}(1 + e^{i\pi}) = 0 \tag{19.3.26}$$

This means that X does not get scattered to -X implying that the gap will close.

19.4 The tight-binding model: electrons in periodic potentials

The largest short-coming of the nearly free electron approximation is that it takes the electrons in a solid to be weakly bound to the ion lattice. We need a complementary method that treats strong potentials $V(\mathbf{r})$, the tight-binding model.

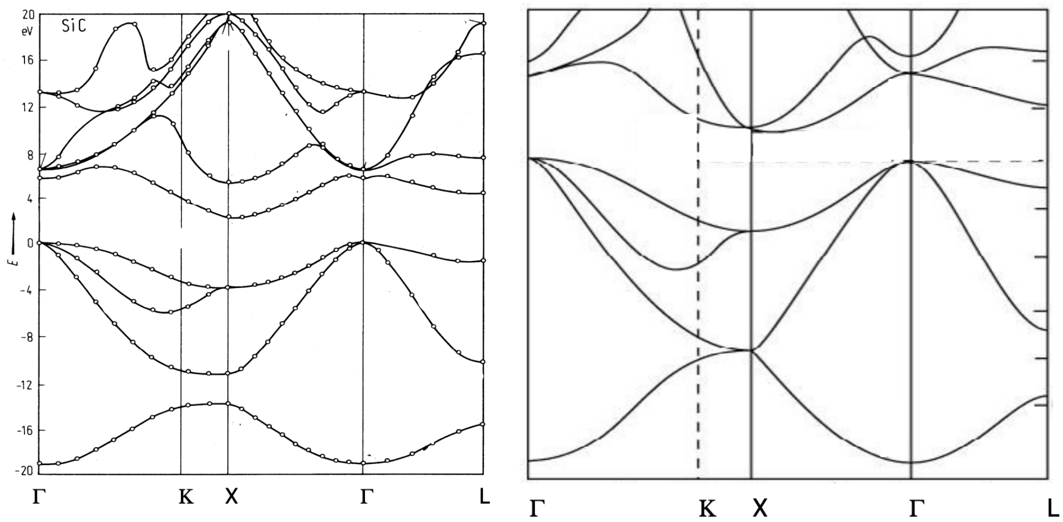


Figure 19.3. The dispersion relation of Silicon carbide (left) and pure Silicon (right).

Let's take a set of ion cores

The associated Hamiltonian is

$$H = -\frac{\hbar^2}{2m}\nabla^2 + \sum_i V(\mathbf{r}_i - \mathbf{R}_i) \tag{19.4.1}$$

Take $\phi_n(\mathbf{r})$ to be the nth single-particle eigenstate centered at an ion at the origin, so that

$$-\frac{\hbar^2}{2m}\nabla^2\phi_n + V(\mathbf{r})\phi_n = E_n\phi_n \tag{19.4.2}$$

We see that the wave-functions centered at two neighbouring sites will have a non-vanishing overlap that leads to their hybridisation. The general wave-function can therefore be taken as a linear combination of atomic orbitals, a superposition of ϕ_n's.

Define the ladder operators $c_n^\dagger(\mathbf{r}_i)$ and $c_n(\mathbf{r}_i)$ which create and annihilate an electron in the orbital $\phi_n(\mathbf{r} - \mathbf{R}_i)$ centered at the ith atom. Mathematically

$$c^\dagger(\mathbf{r}_i) = \tag{19.4.3}$$

Conductors, insulators and semiconductors

20

21 Semi-classical Transport theory

Linear response theory

22

Paramagnetism and Diamagnetism

23

24 Anti-Ferro(i)magnetism and Mean field theory

Phase transitions and Landau theory

25

26

BECs and superfluidity

Superconductivity

27

Part IV

Atomic physics and quantum optics

Part V

Condensed matter field theory

Acknowledgments

This is the most common positions for acknowledgments. A macro is available to maintain the same layout and spelling of the heading.

Note added. This is also a good position for notes added after the paper has been written.

Bibliography

[1] Author, *Title*, *J. Abbrev.* **vol** (year) pg.

[2] Author, *Title*, arxiv:1234.5678.

[3] Author, *Title*, Publisher (year).