Electromagnetic Theory II

Lectures by Tom DeGrand, Spring 2019

Transcribed by Daniel Spiegel

Contents

1 Working in CGS

In this course, we will use cgs/Gaussian units exclusively. There are several advantages of this system. The equations are usually simpler, and there aren't any pesky ε_0's or μ_0's floating around. Most importantly though, the electric and magnetic fields have the same units, as do the scalar and vector potentials. This will be especially nice when we come to special relativity. Below we record the key equations of electromagnetism as they appear in cgs units.

MAXWELL'S EQUATIONS

$$\text{(Gauss)} \quad \boldsymbol{\nabla} \cdot \mathbf{E} = 4\pi\rho \qquad \text{(Faraday)} \quad \boldsymbol{\nabla} \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0$$

$$\text{(No Mag. Monopoles)} \quad \boldsymbol{\nabla} \cdot \mathbf{B} = 0 \qquad \text{(Ampere)} \quad \boldsymbol{\nabla} \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{J}$$

LORENTZ FORCE LAW

$$\mathbf{F} = q\Big(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B}\Big)$$

POTENTIALS

$$\mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A} \qquad \mathbf{E} = -\boldsymbol{\nabla}\Phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$$

GAUGE TRANSFORMATIONS

$$\mathbf{A} \to \mathbf{A} + \boldsymbol{\nabla}\Lambda \qquad \Phi \to \Phi - \frac{1}{c}\frac{\partial \Lambda}{\partial t}$$

LORENZ GAUGE

$$\boldsymbol{\nabla} \cdot \mathbf{A} + \frac{1}{c}\frac{\partial \Phi}{\partial t} = 0$$

POYNTING VECTOR

$$\mathbf{S} \equiv \frac{c}{4\pi}(\mathbf{E} \times \mathbf{B})$$

Let us look at the relation between the electric and magnetic fields in cgs units. In the absence of sources, the fields obey homogeneous wave equations

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\begin{bmatrix}\mathbf{E} \\ \mathbf{B}\end{bmatrix} = 0.$$

These have plane wave solutions, e.g. for a given frequency ω the $\mathbf{E}$ field is

$$\mathbf{E} = \hat{\varepsilon} E_0 e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)},$$

where the wave equation demands $k^2 = \omega^2/c^2$. In addition, Gauss's law $\boldsymbol{\nabla}\cdot\mathbf{E} = 0$ enforces $\mathbf{k}\cdot\hat{\varepsilon} = 0$. The magnetic field has the same exponential dependence.

To relate the $\mathbf{E}$ and $\mathbf{B}$ fields, we look to Faraday's law:

$$-\frac{1}{c}\frac{\partial B}{\partial t} = \boldsymbol{\nabla}\times\mathbf{E} \quad\Longrightarrow\quad \frac{i\omega}{c}\mathbf{B} = i\mathbf{k}\times\mathbf{E}.$$

Since $k = \omega/c$, we get

$$\mathbf{B} = \hat{\mathbf{k}}\times\mathbf{E}$$

Behold the power of cgs units: $\mathbf{E}$ and $\mathbf{B}$ have the same magnitude! Of course, we observe that $\mathbf{k}, \mathbf{E}$, and $\mathbf{B}$ are mutually orthogonal as well.

2 Radiation

We are interested in finding what the electric and magnetic fields look like at different distance regimes from different source configurations. For simplicity and practicality, we will assume there are no troublesome boundaries, dielectrics, or magnetic materials. In effect, $\epsilon = \epsilon_0 = 1$ and $\mu = \mu_0 = 1$.

2.1 Green's Functions for the Wave Equation

In Lorenz gauge, substituting out $\mathbf{E}$ and $\mathbf{B}$ for the potentials in Gauss and Ampere's laws gives us decoupled wave equations for Φ and $\mathbf{A}$.

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\begin{bmatrix}\mathbf{A}\\ \Phi\end{bmatrix} = -\frac{4\pi}{c}\begin{bmatrix}\mathbf{J}\\ c\rho\end{bmatrix} \tag{1}$$

Let's focus on $\mathbf{A}$ for now. In general the program will be to find the potentials, then compute $\mathbf{E}$ and $\mathbf{B}$, then compute $\mathbf{S}$, then interpret the results. To solve the wave equation, we introduce a Green's function satisfying

$$\left(\boldsymbol{\nabla}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)G(\mathbf{x}, t; \mathbf{x}', t') = -4\pi\delta(\mathbf{x}-\mathbf{x}')\delta(t-t'). \tag{2}$$

Physically, the Green's function is the response to a point source introduced at a single instant in time. Once we have the Green's function, we may compute $\mathbf{A}$ as

$$\mathbf{A}(\mathbf{x}, t) = \frac{1}{c}\int d^3x' dt' G(\mathbf{x}, t; \mathbf{x}', t')\mathbf{J}(\mathbf{x}', t'). \tag{3}$$

That this satisfies the wave equation can be easily checked.

It is simpler to find the Green's function in Fourier space. Our conventions for Fourier transforms are

$$f(\mathbf{x},t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\mathbf{x},\omega)e^{-i\omega t}d\omega \tag{4}$$

$$f(\mathbf{x},\omega) = \int_{-\infty}^{\infty} f(\mathbf{x},t)e^{i\omega t}\,dt\,. \tag{5}$$

Fourier transforming the wave equation yields the general form

$$\big(\boldsymbol{\nabla}^2 + k^2\big)\psi(\mathbf{x},\omega) = -4\pi f(\mathbf{x},\omega), \qquad k = \frac{\omega}{c}, \tag{6}$$

known as the ***inhomogeneous Helmholtz equation***. The Green's functions for the Helmholtz equation obey

$$\big(\boldsymbol{\nabla}^2 + k^2\big)G(\mathbf{x},\omega;\mathbf{x}',\omega') = -4\pi\delta^3(\mathbf{x}-\mathbf{x}')(2\pi)\delta(\omega+\omega'). \tag{7}$$

We may partially separate variables as $G(\mathbf{x},\omega;\mathbf{x}',\omega') = G_k(\mathbf{x},\mathbf{x}')(2\pi)\delta(\omega+\omega')$ and solve for G_k, which satisfies

$$\big(\boldsymbol{\nabla}^2 + k^2\big)G_k(\mathbf{x},\mathbf{x}') = -4\pi\delta(\mathbf{x}-\mathbf{x}') \tag{8}$$

Imagine placing $\mathbf{x}'$ at the origin and use spherical coordinates where $R = |\mathbf{x}-\mathbf{x}'|$. The delta function source depends only on R, and therefore so must G_k. With the Laplacian in spherical coordinates, the Helmholtz equation becomes

$$\frac{1}{R}\frac{d^2}{dR^2}(RG_k) + k^2 G_k = -4\pi\delta(R). \tag{9}$$

If $R \neq 0$, then the delta function disappears and we get

$$\frac{d^2}{dR^2}(RG_k) + k^2 RG_k = 0. \tag{10}$$

This has solutions given by

$$G_k(R) = \frac{Ae^{ikR} + Be^{-ikR}}{R}. \tag{11}$$

The delta function turns on as $R \to 0$, but $ikR \to 0$. Furthermore, as $R \to 0$, the Helmholtz equation reduces to Laplace's equation, so we know the solution must be

$$G_k(R) \to \frac{1}{R} \;\text{ as } R \to 0.$$

This requires that

$$A + B = 1.$$

There are still two linearly independent solutions to the Helmholtz equation, given by setting $A = 1$ and $B = 1$ respectively:

$$G_k^{(\pm)}(R) = \frac{e^{\pm ikR}}{R}.$$

Fourier transforming back to a time coordinate, we obtain the Green's function

$$\begin{aligned} G^{(\pm)}(\mathbf{x}, t; \mathbf{x}', t') &= \frac{1}{(2\pi)^2} \iint d\omega\, d\omega'\, G_k^{(\pm)}(R)(2\pi)\delta(\omega + \omega')e^{-i\omega t}e^{-i\omega' t'} && (12) \\ &= \frac{1}{2\pi} \int d\omega\, \frac{e^{\pm ikR}}{R} e^{-i\omega(t-t')} && (13) \end{aligned}$$

With $k = \omega/c$, the exponentials combine and the integral gives a simple delta function:

$$G^{(\pm)}(\mathbf{x}, t; \mathbf{x}', t') = \frac{1}{R}\delta\left(t' - \left(t \mp \frac{R}{c}\right)\right) \tag{14}$$

We see that G^+—the ***retarded Green's function***—exhibits a response at time t to a signal that appeared at an earlier time $t' = t - R/c$, accounting for the propagation speed of light. Conversely, the ***advanced Green's function*** G^- responds to signals in the future; this isn't physical so we won't deal with it.

The formal solution to the wave equation is then

$$\begin{aligned} \mathbf{A}(\mathbf{x}, t) &= \frac{1}{c} \int d^3x'\, dt'\, G^{(+)}(\mathbf{x}, t; \mathbf{x}', t')\mathbf{J}(\mathbf{x}', t') && (15) \\ &= \frac{1}{c} \int d^3x'\, \frac{1}{|\mathbf{x} - \mathbf{x}'|}\mathbf{J}\left(\mathbf{x}', t - \frac{|\mathbf{x} - \mathbf{x}'|}{c}\right) && (16) \end{aligned}$$

Unfortunately, this integral will be prohibitively difficult, and we will be forced to make approximations.

Let's imagine a sinusoidally oscillating source:

$$\begin{aligned} \mathbf{J}(\mathbf{x}, t) &= \mathbf{J}(\mathbf{x})e^{-i\omega t} && (17) \\ \rho(\mathbf{x}, t) &= \rho(\mathbf{x})e^{-i\omega t}. && (18) \end{aligned}$$

This will result in a sinusoidally oscillating response

$$\mathbf{A}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x})e^{-i\omega t}, \tag{19}$$

with

$$\boxed{\mathbf{A}(\mathbf{x}) = \frac{1}{c} \int d^3x'\, \frac{\mathbf{J}(\mathbf{x}')}{R} e^{ikR},} \tag{20}$$

where again $R = |\mathbf{x} - \mathbf{x}'|$.

Do we need to go back and do the whole thing again to get Φ? No. We can get $\mathbf{B}$ straight from $\mathbf{A}$. If we're away from the sources, we can then also get the electric field straight from Ampere's law and the fact that $\mathbf{E}$ will be similarly sinusoidal:

$$\boldsymbol{\nabla} \times \mathbf{B} = \frac{1}{c}\frac{\partial E}{\partial t} = -\frac{i\omega}{c}\mathbf{E}. \tag{21}$$

If we really wanted Φ for whatever reason, we could also just obtain it from the Lorenz gauge condition.

2.2 The Multipole Expansion

Continuing from (20), we will have to make even more approximations. These approximations will be based on comparison of the following distance scales:

$$\begin{aligned} d &= \text{size of the source}, \\ r &= \text{distance to receiver}, \\ \lambda &= \text{wavelength of radiation} = \frac{2\pi}{k}. \end{aligned}$$

In the ***far field approximation*** we work in the ***radiation zone*** $r \gg \lambda$. In this case we'll find $E \sim r^{-1}$ rather than r^{-2}. We can also consider the ***long wavelength limit*** $\lambda \gg d$. This applies to atoms, where typically $d \sim 1\,\mathring{\text{A}}$ and $\lambda \sim 1000\,\mathring{\text{A}}$. In this limit we have

$$c \gg \frac{cd}{\lambda} = \nu d \sim v.$$

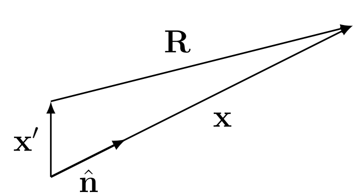
Thus, in the long wavelength limit our sources oscillate at speeds much less than the speed of light. We will often make both the far field and long wavelength approximations.

We can also consider the near zone $\lambda \gg r \gg d$, in which $e^{ikR} \sim 1$. We see from (20) that in this case the potential reduces to the familiar electrostatic potential! A good example of when this approximation is relevant is an avalanche beacon. If a backcountry skier carrying an avalanche beacon is buried alive by an avalanche, the beacon will emit radio waves of $\nu = 457\,\text{kHz}$, $\lambda = 656\,\text{m}$, which will be picked up by a search party when within a few meters of the skier.

Going back to the radiation zone, let's begin utilizing our approximations $kr \gg 1$ and $kd \ll 1$ to obtain a nicer expression for

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c}\int d^3x'\,\frac{\mathbf{J}(\mathbf{x})}{R}e^{ikR}. \tag{20}$$

Let $\hat{\mathbf{n}}$ be a unit vector in the direction of $\mathbf{x}$; see the picture below.



Then $R = |\mathbf{x} - \mathbf{x}'|$ may be approximated as

$$|\mathbf{x} - \mathbf{x}'| = \sqrt{r^2 + r'^2 - 2\mathbf{x}\cdot\mathbf{x}'} \simeq r - \frac{\mathbf{x}\cdot\mathbf{x}'}{r}. = r - \hat{\mathbf{n}}\cdot\mathbf{x}', \tag{22}$$

where $r = |\mathbf{x}|$ and we have approximated to first order in $r' = |\mathbf{x}'|$. In fact, if we keep only the leading behavior in $(kr)^{-1}$, then the R^{-1} in (20) may be simply approximated by $R \simeq r$. For the exponential e^{ikR} we must keep $R \simeq r - \hat{\mathbf{n}}\cdot\mathbf{x}'$ since the small parameter there is kr', not kr. Plugging this into (20) yields

$$\mathbf{A}(\mathbf{x}) = \frac{e^{ikr}}{cr}\int d^3x'\, e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}\mathbf{J}(\mathbf{x}'). \tag{23}$$

Note that the prefactor e^{ikr}/cr is just an outgoing spherical wave. If we wanted to take the curl to get $\mathbf{B}$, we would have to apply it to e^{ikr}, r^{-1}, and $e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}$, which sounds painful. But applying the curl to r^{-1} would yield r^{-2}, which is subleading in the radiation zone! This simplifies things.

In the integral, the $\mathbf{x}'$ is bounded by the scale d. Since $k \sim \lambda^{-1}$ and we're in the long wavelength limit, $k\hat{\mathbf{n}}\cdot\mathbf{x}'$ is small, so we can expand the exponential. This yields

$$\boxed{\mathbf{A}(\mathbf{x}) = \frac{e^{ikr}}{cr}\sum_{\ell}\frac{1}{\ell!}\int d^3x'\, (-ik\hat{\mathbf{n}}\cdot\mathbf{x}')^{\ell}\mathbf{J}(\mathbf{x}')} \tag{24}$$

This is the ***multipole expansion***, and is significanlty more useable than (20). The terms of the series go like $(kd)^{\ell}$, which converges very fast in the long wavelength limit. Thus, one typically stops computing at the first nonzero term.

Electric Monopoles

Before diving into the multipole expansion, let's take a moment to look at electric monopole radiation, which is not included in (24). The electric monopole is the lowest order term in $(kr)^{-1}$ in the scalar potential

$$\Phi(\mathbf{x}, t) = \iint d^3x'\, dt'\, G^{(+)}(\mathbf{x}, t; \mathbf{x}', t')\rho(\mathbf{x}', t') = \int d^3x'\, \rho\left(\mathbf{x}', t - \frac{R}{c}\right)\frac{1}{R}. \tag{25}$$

We approximate $R^{-1} \sim r^{-1}$ and obtain

$$\Phi_0(\mathbf{x}, t) = \frac{1}{r}\int d^3x'\, \rho\left(\mathbf{x}', t - \frac{R}{c}\right) = \frac{Q(t - R/c)}{r} = \frac{Q}{r}, \tag{26}$$

the last line following by charge conservation. This is just an electrostatic potential. The corresponding electric field goes like r^{-2}, which we neglect in the radiation zone.

Electric Dipole Radiation

For something more interesting, we consider electric dipole radiation. This corresponds to the $\ell = 0$ term in the multipole expansion:

$$\mathbf{A}(\mathbf{x}) = \frac{e^{ikr}}{cr}\int d^3x'\,\mathbf{J}(\mathbf{x}') \tag{27}$$

It's sometimes easier to work with ρ than $\mathbf{J}$, so let's try to switch variables. We can use the continuity equation

$$\boldsymbol{\nabla}\cdot\mathbf{J} = -\frac{\partial\rho}{\partial t} = i\omega\rho. \tag{28}$$

Looking at a single component of the current integral, we have

$$\int d^3x'\,J_i(\mathbf{x}') = \int d^3x'\,(\boldsymbol{\nabla}'x_i')\cdot\mathbf{J} = \int d^3x'\,\boldsymbol{\nabla}'\cdot(x_i'\mathbf{J}) - \int d^3x\,x_i'\boldsymbol{\nabla}'\cdot\mathbf{J}. \tag{29}$$

The first term is a surface integral; letting the surface be larger than the range of the current puts that term to zero. Applying the continuity equation to the second term yields

$$\int d^3x'\,\mathbf{J} = -i\omega\int d^3x'\,\rho(\mathbf{x}')\mathbf{x}' = -i\omega\mathbf{p}, \tag{30}$$

where $\mathbf{p}$ is the electric dipole moment. Thus,

$$\boxed{\mathbf{A}(\mathbf{x}) = -ik\mathbf{p}\frac{e^{ikr}}{r}.} \tag{31}$$

Let's get $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{S}$ for electric dipole radiation.

$$\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} = -ik\boldsymbol{\nabla}\times\left(\frac{e^{ikr}}{r}\mathbf{p}\right) = ik\mathbf{p}\times\boldsymbol{\nabla}\left(\frac{e^{ikr}}{r}\right) \tag{32}$$

The derivative of r^{-1} may be neglected in the radiation zone, and

$$\boldsymbol{\nabla}r = \frac{1}{r}\sum_i x_i\hat{\mathbf{x}}_i = \hat{\mathbf{n}}, \tag{33}$$

so the chain rule gives

$$\boxed{\mathbf{B} = k^2\frac{e^{ikr}}{r}[\hat{\mathbf{n}}\times\mathbf{p}].} \tag{34}$$

As a sanity check, note that $\mathbf{B}$ is a transverse wave.

For $\mathbf{E}$, we compute

$$\begin{aligned}\mathbf{E} &= \frac{i}{k}\boldsymbol{\nabla}\times\mathbf{B} &(35)\\ &= \frac{ik}{r}\left[\boldsymbol{\nabla}\left(e^{ikr}\right)\times(\hat{\mathbf{n}}\times\mathbf{p}) + e^{ikr}\boldsymbol{\nabla}\times(\hat{\mathbf{n}}\times\mathbf{p})\right] &(36)\\ &= ik\frac{e^{ikr}}{r}[(ik)\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{p}) + (\mathbf{p}\cdot\boldsymbol{\nabla})\hat{\mathbf{n}} - \mathbf{p}(\boldsymbol{\nabla}\cdot\hat{\mathbf{n}})] &(37)\end{aligned}$$

However, observe that

$$\frac{\partial}{\partial x^i}\hat{\mathbf{n}} = \frac{1}{r}\hat{\mathbf{x}}_i - \frac{x_i}{r^2}\hat{\mathbf{n}} \quad \text{and} \quad \boldsymbol{\nabla}\cdot\hat{\mathbf{n}} = \frac{2}{r}; \tag{38}$$

both of these terms are subleading in the radiation zone. Thus,

$$\boxed{\mathbf{E} = -k^2\frac{e^{ikr}}{r}[\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{p})] = k^2\frac{e^{ikr}}{r}[\mathbf{p}-(\hat{\mathbf{n}}\cdot\mathbf{p})\hat{\mathbf{n}}].} \tag{39}$$

We note that

$$\mathbf{E} = -\hat{\mathbf{n}}\times\mathbf{B}, \tag{40}$$

a characteristic of the radiation zone.

For the time averaged Poynting vector, we have

$$\langle\mathbf{S}\rangle = \frac{c}{8\pi}\,\mathrm{Re}\,(\mathbf{E}\times\mathbf{B}^*) \tag{41}$$

$$= -\frac{c}{8\pi}\frac{k^4}{r^2}\{[(\hat{\mathbf{n}}\cdot\mathbf{p})\hat{\mathbf{n}}-\mathbf{p}]\times(\hat{\mathbf{n}}\times\mathbf{p})\} \tag{42}$$

$$= -\frac{c}{8\pi}\frac{k^4}{r^2}\left[(\hat{\mathbf{n}}\cdot\mathbf{p})^2\hat{\mathbf{n}} - p^2\hat{\mathbf{n}}\right] \tag{43}$$

and the power radiated per solid angle is

$$\boxed{\frac{dP}{d\Omega} = r^2\hat{\mathbf{n}}\cdot\langle\mathbf{S}\rangle = \frac{ck^4}{8\pi}\left[p^2-(\hat{\mathbf{n}}\cdot\mathbf{p})^2\right].} \tag{44}$$

Alternatively, we may express the power from any transverse wave in terms of the electric field as

$$\frac{dP}{d\Omega} = \frac{r^2c}{8\pi}\hat{\mathbf{n}}\cdot(\mathbf{E}\times\mathbf{B}^*) = \frac{cr^2}{8\pi}\mathbf{E}\cdot(\mathbf{B}^*\times\hat{\mathbf{n}}) = \frac{cr^2}{8\pi}|\mathbf{E}|^2. \tag{45}$$

Thanks to Youssef Hassan for pointing this out. The power is constant in r, which allows us to see things far away, and goes like ω^4, so higher frequency light radiates with vastly greater power. This is one factor that goes into making the sky blue.

While (44) gives the total power radiated per unit solid angle, for a particular problem we might be more interested in the power radiated in a certain polarization of $\mathbf{E}$. Recall that $\mathbf{E}$ can be decomposed as

$$\mathbf{E} = \sum_\lambda \hat{\boldsymbol{\epsilon}}_\lambda(\hat{\boldsymbol{\epsilon}}_\lambda^*\cdot\mathbf{E}),$$

so that the component of $\mathbf{E}$ in the polarization $\hat{\boldsymbol{\epsilon}}_\lambda$ is $\hat{\boldsymbol{\epsilon}}_\lambda^*\cdot\mathbf{E}$ (see Jackson Section 7.2). To obtain the power radiated in this polarization, we simply replace $|\mathbf{E}|^2$ in (45) with the value square of this component.

$$\frac{dP}{d\Omega}(\hat{\boldsymbol{\epsilon}}_\lambda) = \frac{cr^2}{8\pi}|\hat{\boldsymbol{\epsilon}}_\lambda^*\cdot\mathbf{E}|^2 \tag{46}$$

Example 1. Suppose $\mathbf{p}$ is fixed in some direction, i.e. $\mathbf{p} = p\hat{\mathbf{z}}$ (with harmonic time dependence). Then

$$\frac{dP}{d\Omega} = \frac{ck^4}{8\pi} p^2 \sin^2\theta. \tag{47}$$

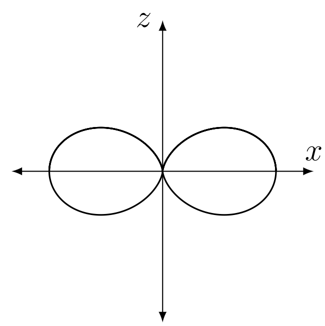
For the total power, we need the integral

$$\int_{-1}^{1} \left[1 - \cos^2\theta\right] d(\cos\theta) = \frac{4}{3}. \tag{48}$$

Integrating over all solid angles then yields

$$P = \frac{1}{3} ck^4 p^2. \tag{49}$$

The antenna pattern for the dipole is displayed below.

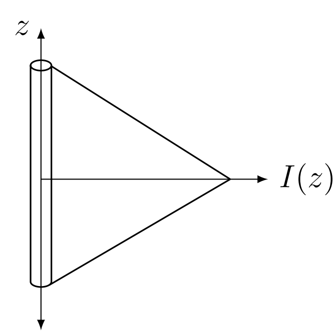


Example 2. Consider a short antenna of length d centered on the origin oriented along the z axis. We assume a harmonically varying current

$$I = I(z) e^{-i\omega t}, \tag{50}$$

and require $I(\pm d/2) = 0$ so that charge doesn't fall off the ends of the antenna. For concreteness, let's take

$$I(z) = I_0 \left[1 - \frac{2|z|}{d}\right],$$



We can obtain the charge density λ with the continuity equation

$$i\omega\lambda = \frac{dI}{dz} \quad \Longrightarrow \quad \lambda = \pm\frac{2I_0 i}{\omega d}, \tag{51}$$

where the plus sign applies for $z > 0$ and the minus sign applies for $z < 0$. The dipole moment is

$$p_z = \int_{-d/2}^{d/2} z\lambda(z)\, dz = \frac{iI_0 d}{2\omega} \tag{52}$$

The antenna pattern is then

$$\frac{dP}{d\Omega} = \frac{ck^4}{8\pi}\left(\frac{I_0 d}{2ck}\right)^2 \sin^2\theta = \frac{I_0^2 (kd)^2}{32\pi c}\sin^2\theta, \tag{53}$$

with total power

$$P = \frac{I_0^2 (kd)^2}{12c}. \tag{54}$$

The total power goes like I_0^2, just like with Joule's law for ordinary circuits. With a 1/2 from time averaging, we define the ***radiation resistance*** R by

$$P = \frac{1}{2} I_0^2 R. \tag{55}$$

For this problem, we have $R = (kd)^2/6c$. In Ohms, this is

$$R(\text{Ohms}) = 30cR_{cgs} = 5(kd)^2\,\Omega. \tag{56}$$

For a half wave antenna $kd = 2\pi d/\lambda = \pi$, so

$$R(\text{Ohms}) = 50\,\Omega. \tag{57}$$

This is the impedance of a standard coaxial cable.

Magnetic Dipole and Electric Quadrupole Radiation

Both the magnetic dipole and electric quadrupole radiation come from the $\ell = 1$ term in the multipole expansion

$$\mathbf{A}(\mathbf{x}) = -ik\frac{e^{ikr}}{cr}\int d^3x'\, (\hat{\mathbf{n}}\cdot\mathbf{x}')\mathbf{J}(\mathbf{x}') \tag{58}$$

Recall that we can write the current as $\mathbf{J} = \mathbf{J}_t + \mathbf{J}_\ell$, where t stands for transverse, ℓ stands for longitudinal, and

$$\boldsymbol{\nabla}\cdot\mathbf{J}_\ell = -\frac{\partial\rho}{\partial t}, \qquad \nabla\cdot\mathbf{J}_t = 0. \tag{59}$$

Intuitively, the electric quadrupole radiation comes from $\mathbf{J}_\ell$ and the magnetic dipole radiation comes from $\mathbf{J}_t$.

Let's work on the integral in (58) with the BAC-CAB rule

$$\hat{\mathbf{n}}\times(\mathbf{x}'\times\mathbf{J}) = \mathbf{x}'(\hat{\mathbf{n}}\cdot\mathbf{J}) - \mathbf{J}(\hat{\mathbf{n}}\cdot\mathbf{x}'). \tag{60}$$

The second term on the right is our integrand. Manipulating this further we write

$$(\hat{\mathbf{n}}\cdot\mathbf{x}')\mathbf{J} = \frac{1}{2}[(\hat{\mathbf{n}}\cdot\mathbf{x}')\mathbf{J} + (\hat{\mathbf{n}}\cdot\mathbf{J})\mathbf{x}'] - \frac{1}{2}\hat{\mathbf{n}}\times(\mathbf{x}'\times\mathbf{J}). \tag{61}$$

The term in square brackets will give us the electric quadrupole and the cross products will give us the magnetic dipole moment.

Let's consider the magnetic dipole first. The magnetic dipole moment was defined as

$$\mathbf{m} = \int d^3x\,\mathbf{M} = \frac{1}{2c}\int d^3x\,(\mathbf{x}\times\mathbf{J}). \tag{62}$$

Plugging (61) into (58) and substituting in (62) yields

$$\boxed{\mathbf{A}_{\mathrm{M1}}(\mathbf{x}) = ik(\hat{\mathbf{n}}\times\mathbf{m})\frac{e^{ikr}}{r}}$$

Taking a curl yields the magnetic field

$$\mathbf{B} = \boldsymbol{\nabla}\times\mathbf{A} = ik\hat{\mathbf{n}}\times\mathbf{A} = \boxed{-k^2\frac{e^{ikr}}{r}\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{m}),} \tag{63}$$

in a very similar fashion that the electric dipole $\mathbf{E}$ was obtained from the electric dipole $\mathbf{B}$. The electric field is

$$\mathbf{E} = \frac{i}{k}\boldsymbol{\nabla}\times\mathbf{B} = -\hat{\mathbf{n}}\times\mathbf{B} = \boxed{-k^2\frac{e^{ikr}}{r}(\hat{\mathbf{n}}\times\mathbf{m}).} \tag{64}$$

We emphasize the similarities between the electric and magnetic dipoles in the table below.

	E1	M1
$\mathbf{B}$	$\propto \hat{\mathbf{n}}\times\mathbf{p}$	$\propto -\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{m})$
$\mathbf{E}$	$\propto -\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{p})$	$\propto -\hat{\mathbf{n}}\times\mathbf{m}$
$dP/d\Omega$	$(ck^4/8\pi)\lvert\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{p})\rvert^2$	$(ck^4/8\pi)\lvert\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{m})\rvert^2$
P	$ck^4p^2/3$	$ck^4m^2/3$

The proportionality constant for $\mathbf{E}$ and $\mathbf{B}$ is always k^2e^{ikr}/r.

Now let's consider the electric quadrupole:

$$\mathbf{A}_{\mathrm{E2}}(\mathbf{x}) = -\frac{ik}{2c}\frac{e^{ikr}}{r}\int d^3x'\,[(\hat{\mathbf{n}}\cdot\mathbf{x}')\mathbf{J} + \mathbf{x}'(\hat{\mathbf{n}}\cdot\mathbf{J})]. \tag{65}$$

Again, The goal is to express this in terms of ρ. The ith component of the first term can be written as

$$\int d^3x'\,(\hat{\mathbf{n}}\cdot\mathbf{x}')J_i = \int d^3x\,(\hat{\mathbf{n}}\cdot\mathbf{x}')[(\boldsymbol{\nabla}'x_i')\cdot\mathbf{J}] \tag{66}$$

$$= \int d^3x'\,\boldsymbol{\nabla}'\cdot[(\hat{\mathbf{n}}\cdot\mathbf{x}')x_i'\mathbf{J}] - \int d^3x'\,x_i'\boldsymbol{\nabla}'\cdot[(\hat{\mathbf{n}}\cdot\mathbf{x}')\mathbf{J}]. \tag{67}$$

The first term is a surface integral, so it dies. For the second term we use the vector identity

$$\boldsymbol{\nabla}' \cdot [(\hat{\mathbf{n}} \cdot \mathbf{x}')\mathbf{J}] = [\boldsymbol{\nabla}'(\hat{\mathbf{n}} \cdot \mathbf{x}')] \cdot \mathbf{J} + (\hat{\mathbf{n}} \cdot \mathbf{x}')\boldsymbol{\nabla}' \cdot \mathbf{J} = \hat{\mathbf{n}} \cdot \mathbf{J} + (\hat{\mathbf{n}} \cdot \mathbf{x}')\boldsymbol{\nabla}' \cdot \mathbf{J} \tag{68}$$

Plugging (68) into (67) and (67) into (65), the first term of (68) cancels with the second term of (65) and we're left with

$$\mathbf{A}_{\mathrm{E2}} = \frac{ik}{2c}\frac{e^{ikr}}{r}\int d^3x'\,\mathbf{x}'(\hat{\mathbf{n}} \cdot \mathbf{x}')\boldsymbol{\nabla} \cdot \mathbf{J} = -\frac{k^2}{2}\frac{e^{ikr}}{r}\int d^3x'\,\mathbf{x}'(\hat{\mathbf{n}} \cdot \mathbf{x}')\rho \tag{69}$$

Note that the integrand has two powers of $\mathbf{x}'$, unlike the dipole moment. There's an extra factor of k to compensate. We take a curl to get

$$\mathbf{B}_{\mathrm{E2}} = ik\hat{\mathbf{n}} \times \mathbf{A}_{\mathrm{E2}} = -\frac{ik^3}{2}\frac{e^{ikr}}{r}\int d^3x'\,(\hat{\mathbf{n}} \times \mathbf{x}')(\hat{\mathbf{n}} \cdot \mathbf{x}')\rho \tag{70}$$

We expressed dipole fields in terms of the dipole moment; we can express the quadrupole fields in terms of the quadrupole moment tensor. Recall the definition

$$Q_{ij} = \int d^3x'\,\rho(\mathbf{x}')\left[3x'_i x'_j - \delta_{ij}x'^2\right]. \tag{71}$$

Note that the quadrupole moment tensor is symmetric and traceless. We can act on a vector with Q to get the "quadrupole vector",

$$q_i(\hat{\mathbf{n}}) = Q_{ij}\hat{n}_j \tag{72}$$

We can compute the cross product

$$\hat{\mathbf{n}} \times \mathbf{q}(\hat{\mathbf{n}}) = \int d^3x'\,\left[3(\hat{\mathbf{n}} \times \mathbf{x}')(\hat{\mathbf{n}} \cdot \mathbf{x}') - x'^2\hat{\mathbf{n}} \times \hat{\mathbf{n}}\right]\rho = 3\int d^3x'\,(\hat{\mathbf{n}} \times \mathbf{x}')(\hat{\mathbf{n}} \cdot \mathbf{x}'), \tag{73}$$

which looks a lot like the magnetic field! We rewrite (70) as

$$\mathbf{B}_{E2} = -\frac{ik^3}{6}\frac{e^{ikr}}{r}\hat{\mathbf{n}} \times \mathbf{q}(\hat{\mathbf{n}}). \tag{74}$$

The electric field is then obtained as

$$\mathbf{E}_{E2} = -\hat{\mathbf{n}} \times \mathbf{B}_{E2} = \frac{ik^3}{6}\frac{e^{ikr}}{r}(\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \mathbf{q}(\hat{\mathbf{n}}))). \tag{75}$$

This is just like dipole radiation except we have $\mathbf{q}(\hat{\mathbf{n}})$ instead of $\mathbf{p}$! There's also an extra minus sign, the factor of 1/6, and an extra factor of k to account for the dimensions of $\mathbf{q}$.

We can steal the equation for the radiation from the dipole equation with the replacement $\mathbf{p} \to k\mathbf{q}(\hat{\mathbf{n}})/6$:

$$\boxed{\frac{dP}{d\Omega} = \frac{ck^6}{288\pi}|\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \mathbf{q}(\hat{\mathbf{n}}))|^2 = \frac{ck^6}{288\pi}\left[|\mathbf{q}|^2 - |\hat{\mathbf{n}} \cdot \mathbf{q}|^2\right].} \tag{76}$$

We will employ a few tricks to get the total power, but first we explicitly write out

$$|\mathbf{q}|^2 = Q^*_{\alpha\beta} Q_{\alpha\gamma} n_\beta n_\gamma, \tag{77}$$

$$|\hat{\mathbf{n}} \cdot \mathbf{q}|^2 = Q^*_{\alpha\beta} Q_{\gamma\delta} n_\alpha n_\beta n_\gamma n_\delta, \tag{78}$$

where summation over repeated indices is implied.

The trick is in doing the integral over the components of $\hat{\mathbf{n}}$. Observe

$$\int n_\alpha n_\beta d\Omega = \frac{1}{3}\delta_{\alpha\beta}\left(\int d\Omega \sum_\gamma n_\gamma^2\right) = \frac{4\pi}{3}\delta_{\alpha\beta}. \tag{79}$$

For the integral $\int n_\alpha n_\beta n_\gamma n_\delta \, d\Omega$, observe that if any index value appears only once among $\alpha, \beta, \gamma, \delta$, then the integral will vanish since the integration by symmetry. Thus, we need only consider the cases where there are two distinct pairs of indices and the case where all four indices are the same. By symmetry, we may choose these indices however we please. The four-of-a-kind case is

$$\int n_z^4 \, d\Omega = 2\pi \int_{-1}^{1} \cos^4\theta \, d(\cos\theta) = \frac{4\pi}{5}. \tag{80}$$

The two-pair case then follows:

$$\int n_z^2 n_y^2 \, d\Omega = \int n_y^2 \, d\Omega - \int n_x^2 n_y^2 \, d\Omega - \int n_y^4 \, d\Omega \tag{81}$$

$$\implies \quad \int n_z^2 n_y^2 \, d\Omega = \frac{1}{2}\left[\frac{4\pi}{3} - \frac{4\pi}{5}\right] = \frac{4\pi}{15}. \tag{82}$$

Then we combine (80) and (82) into

$$\int n_\alpha n_\beta n_\gamma n_\delta \, d\Omega = \frac{4\pi}{15}[\delta_{\alpha\beta}\delta_{\gamma\delta} + \delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}] \tag{83}$$

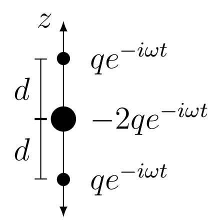
Applying (79) and (83) to integrate $dP/d\Omega$ yields

$$\int d\Omega \left[|\mathbf{q}|^2 - |\hat{\mathbf{n}} \cdot \mathbf{q}|^2\right] = 4\pi\left[\frac{1}{3}\sum_{\alpha\beta}|Q_{\alpha\beta}|^2 - \frac{1}{15}\left(\sum_\alpha Q^*_{\alpha\alpha}\sum_\beta Q_{\beta\beta} + 2\sum_{\alpha\beta}|Q_{\alpha\beta}|^2\right)\right] \tag{84}$$

The middle term is zero since the quadrupole tensor is traceless. Thus, the total power is

$$\boxed{P_{E2} = \frac{ck^6}{360}\sum_{\alpha\beta}|Q_{\alpha\beta}|^2.} \tag{85}$$

Example 3. Consider the configuration of point charges shown below.



It is straightforward to compute the components of the quadrupole tensor:

$$Q = Q_0 \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad Q_0 \equiv 4qd^2. \tag{86}$$

We compute $|\mathbf{q}(\hat{\mathbf{n}})|^2$ and $|\hat{\mathbf{n}} \cdot \mathbf{q}(\hat{\mathbf{n}})|^2$:

$$|\mathbf{q}|^2 = Q_0^2 \begin{bmatrix} n_x & n_y & n_z \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}^2 \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \frac{1}{4} Q_0^2 \big(\sin^2\theta + 4\cos^2\theta\big) \tag{87}$$

$$|\hat{\mathbf{n}} \cdot \mathbf{q}(\hat{\mathbf{n}})|^2 = Q_0^2 \left(\begin{bmatrix} n_x & n_y & n_z \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \right)^2 = \frac{1}{4} Q_0^2 \big(\sin^2\theta - 2\cos^2\theta\big)^2 \tag{88}$$

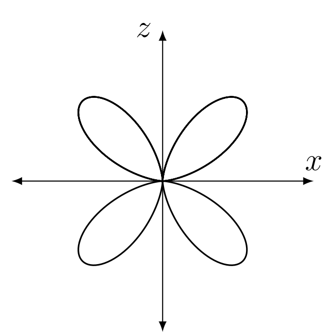
After plugging these into (76) and applying a few trig identities we arrive at

$$\frac{dP}{d\Omega} = \frac{ck^6 Q_0^2}{512} \sin^2 2\theta. \tag{89}$$

The total power radiated can be obtained by integrating the above or by summing the quadrupole components, with the result

$$P = \frac{ck^6 Q_0^2}{240}. \tag{90}$$

The antenna pattern is displayed below. It has twice as many lobes as the dipole antenna pattern.



2.3 Thin Antennas

Consider an antenna of length d along the z axis centered on the origin. In the radiation zone, the vector potential is

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c}\frac{e^{ikr}}{r}\int d^3x'\, e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}\mathbf{J}(\mathbf{x}'). \tag{91}$$

The current must take the form

$$\mathbf{J}(\mathbf{x}') = \hat{\mathbf{z}} I_0 f(z')\delta(x')\delta(y')\theta\left(\frac{d}{2} - |z|\right), \tag{92}$$

where we require $f(-d/2) = f(d/2) = 0$ so that we don't have charge flowing off the antenna. The vector potential takes the form

$$\mathbf{A}(\mathbf{x}) = \hat{\mathbf{z}}\frac{I_0}{c}\frac{e^{ikr}}{r}F(\theta), \tag{93}$$

$$F(\theta) \equiv \int_{-d/2}^{d/2} dz'\, f(z')e^{-ikz'\cos\theta}. \tag{94}$$

From here we can get $\mathbf{B} = ik\hat{\mathbf{n}}\times\mathbf{A}$ and $\mathbf{E} = -\hat{\mathbf{n}}\times\mathbf{B}$ as usual. The antenna pattern will be

$$\frac{dP}{d\Omega} = \frac{cr^2}{8\pi}\hat{\mathbf{n}}\cdot(\mathbf{E}\times\mathbf{B}^*) \tag{95}$$

$$= -\frac{I_0^2k^2}{8\pi c}|F(\theta)|^2\hat{\mathbf{n}}\cdot((\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\hat{\mathbf{z}}))\times(\hat{\mathbf{n}}\times\hat{\mathbf{z}})) \tag{96}$$

$$= \frac{I_0^2k^2}{8\pi c}|F(\theta)|^2\sin^2\theta. \tag{97}$$

The interesting part will come from $|F(\theta)|^2$, which will alter the $\sin^2\theta$ dipole pattern.

For example, let's suppose

$$f(z) = \sin\left(\frac{kd}{2} - k|z|\right). \tag{98}$$

One can explicitly compute $F(\theta)$ with the result

$$F(\theta) = \frac{2}{k}\left[\frac{\cos\left(\frac{kd}{2}\cos\theta\right) - \cos\left(\frac{kd}{2}\right)}{\sin^2\theta}\right]. \tag{99}$$

This is exact. However, in the long wavelength limit we can expand with kd as our small parameter. Expanding each cosine to lowest nontrivial order yields

$$F(\theta) = \frac{2}{k}\frac{1}{2}\left(\frac{kd}{2}\right)^2\frac{[-\cos^2\theta + 1]}{\sin^2\theta} = \frac{(kd)^2}{4k} \tag{100}$$

Thus, in the long-wavelength limit, $|F(\theta)|$ doesn't alter the antenna pattern, and we just have a regular dipole.

What about for a half-wavelength antenna $kd = \pi$? Then $\cos(kd/2) = 0$ and we're left with

$$\frac{dP}{d\Omega} = \frac{I_0^2}{2\pi c} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin^2\theta}, \tag{101}$$

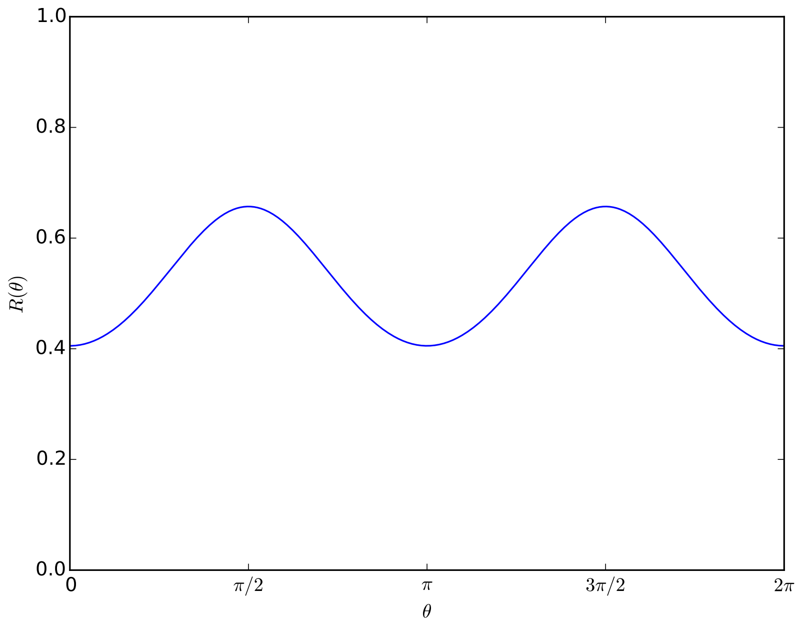
whereas the long-wavelength regular dipole is

$$\left(\frac{dP}{d\Omega}\right)_{\text{dip}} = \frac{I_0^2 k^2}{8\pi c} \frac{\pi^4}{16k^2} \sin^2\theta = \frac{I_0^2 \pi^3}{128c} \sin^2\theta \tag{102}$$

The ratio is

$$R(\theta) = \frac{dP/d\Omega}{(dP/d\Omega)_{\text{dip}}} = \frac{64}{\pi^4} \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin^4\theta}. \tag{103}$$

The plot below shows that the dipole approximation is an overestimate and that the true pattern oscillates about it. Then again, using the dipole approximation may not be valid when $kd = \pi$ is not in the long-wavelength limit.



3 Scattering

The basic picture of scattering is that a plane wave beam interacts with a target, and produces an outgoing wave. You could write this as

$$\psi \sim e^{ikz} + \frac{e^{ikr}}{r} F(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0), \tag{1}$$

where $\hat{\mathbf{n}}_0$ is the direction of the incoming beam, and $\hat{\mathbf{n}}$ is the direction from the target to the detector. Usually one choose $\hat{\mathbf{n}} \neq \hat{\mathbf{n}}_0$ to avoid frying the detector! The intensity of radiation into the detector is

$$I \sim \frac{dP}{d\Omega} \sim r^2 \left| \frac{e^{ikr}}{r} F(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0) \right|^2 = |F(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0)|^2. \tag{2}$$

Typically, F is proportional to the amplitude of the incident beam. In other words, the rate of counts into your detector is proportional to the rate of particles going in. We'll factor out that dependence and write

$$\psi = E_0 e^{ikz} + E_0 \frac{e^{ikr}}{r} f(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0), \tag{3}$$

where $f(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0) = f(\theta)$ is the ***scattering amplitude*** and has no E_0 dependence. Similarly, we'll be more interested in the ***differential cross section***

$$\frac{d\sigma}{d\Omega} \sim \frac{1}{E_0^2} \frac{dP}{d\Omega}. \tag{4}$$

There are typically two paths to obtain $f(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0)$. One is a perturbative approach using the Born approximation, which can be done to arbitrarily high order. In quantum mechanics we would do this if we had a short-range, shallow potential. Alternatively, we can proceed by looking for exact solutions to our differential equations, typically the inhomogeneous Helmholtz equation.

Looking for exact solutions has the benefit of giving exact results. One of these results is the ***Optical theorem***

$$\sigma_{\text{tot}} = \int \frac{d\sigma}{d\Omega} \, d\Omega = \frac{4\pi}{k} \operatorname{Im} f(\hat{\mathbf{n}}, \hat{\mathbf{n}}_0) \tag{5}$$

Another nice exact result comes from expanding the amplitude in terms of Legendre polynomials

$$f(\theta) = \sum_{\ell} P_{\ell}(\cos\theta) \frac{e^{2i\delta_{\ell}(k)} - 1}{2ik} \tag{6}$$

After finding the $\delta_{\ell}(k)$ phase shifts, we get $f(\theta)$ exaclty. This form is typically dealt with in a quantum mechanics class; we won't touch it here.

In general, there is a limit we can take in which things simplify drastically. This is the low energy, long-wavelength limit $kd \ll 1$. Typically, in this limit only one or two terms of the expansion of $f(\theta)$ above matter. In this limit, only the gross features of the scattering target matter, the details not so much.

Let's go back and define the differential cross section precisely. The radiated power distribution is

$$\frac{dP}{d\Omega} = \frac{cr^2}{8\pi} |\mathbf{E}|^2 \propto |E_0|^2 \tag{7}$$

The energy flux of the initial beam is

$$\Phi_i = \frac{c}{8\pi}|E_0|^2, \tag{8}$$

which we see has units of energy/(area · time). We define the differential cross section to be

$$\frac{d\sigma}{d\Omega} \equiv \frac{1}{\Phi_i}\frac{dP}{d\Omega}, \tag{9}$$

which has units of area.

3.1 Scattering off Spheres

Dielectric Sphere

We'll demonstrate the methods of scattering through a few prototypical examples. Imagine we have a dielectric sphere of radius a and permittivity ε at the origin. Let's send in a wave

$$\mathbf{E}_{\text{in}} = \hat{\boldsymbol{\epsilon}}_0 E_0 e^{ik\hat{\mathbf{n}}_0\cdot\mathbf{x}}. \tag{10}$$

As mentioned earlier, things simplify when we take the long-wavelength limit $ka \ll 1$. In this case, for all practical purposes the electric field is uniform over the sphere. The problem of a dielectric sphere in a uniform field was done in the first semester; the sphere acquires an induced dipole moment

$$\mathbf{p} = \left(\frac{\varepsilon - 1}{\varepsilon + 2}\right) a^3 \mathbf{E}_{\text{in}}(t) \tag{11}$$

We know from our study of dipole radiation that this gives scattered electric and magnetic fields

$$\mathbf{B}_{\text{out}} = k^2 \frac{e^{ikr}}{r}\hat{\mathbf{n}} \times \mathbf{p} \tag{12}$$

$$\mathbf{E}_{\text{out}} = -\hat{\mathbf{n}} \times \mathbf{B} \tag{13}$$

The total radiated power distribution is

$$\frac{dP}{d\Omega} = \frac{cr^2}{8\pi}|\mathbf{E}|^2 = \frac{ck^4a^6}{8\pi}\left|\frac{\varepsilon-1}{\varepsilon+2}\right|^2 |E_0|^2 |\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}}_0)|^2 \tag{14}$$

This is familiar, except that now the radiated power explicitly depends on the incoming polarization vector $\hat{\boldsymbol{\epsilon}}_0$. In general, we'll be concerned with the ougtoing polarization vector too. Recall Eq. 46 of Section 2:

$$\frac{dP}{d\Omega}(\hat{\boldsymbol{\epsilon}}_{\text{out}}) = \frac{cr^2}{8\pi}|\hat{\boldsymbol{\epsilon}}^*_{\text{out}} \cdot \mathbf{E}_{\text{out}}|^2. \tag{15}$$

Thus, the radiated power distribution in a specific outgoing polarization $\hat{\boldsymbol{\epsilon}}$ for a specific ingoing polarization $\hat{\boldsymbol{\epsilon}}_0$ is

$$\frac{dP}{d\Omega}(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\epsilon}}_0) = \frac{ck^4 a^6}{8\pi}\left|\frac{\epsilon - 1}{\epsilon + 2}\right|^2 |E_0|^2 |\hat{\boldsymbol{\epsilon}}^* \cdot (\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}}_0))|^2 \tag{16}$$

$$= \frac{ck^4 a^6}{8\pi}\left|\frac{\varepsilon - 1}{\varepsilon + 2}\right|^2 |E_0|^2 |\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2, \tag{17}$$

where the second step follows by $\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\mathbf{n}} = 0$. We'll stick to the convention of $\hat{\boldsymbol{\epsilon}}_0$ being the incoming polarization and $\hat{\boldsymbol{\epsilon}}$ being the outgoing polarization.

Taking the differential cross section yields

$$\frac{d\sigma}{d\Omega} = \frac{1}{\Phi_i}\frac{dP}{d\Omega} = (ka)^4 a^2 \left|\frac{\varepsilon - 1}{\varepsilon + 2}\right|^2 |\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2. \tag{18}$$

The differential cross section has units of area, as it should. We also note that

$$\frac{d\sigma}{d\Omega} \propto k^4 \propto \lambda^{-4}. \tag{19}$$

This is ***Rayleigh's law***; the factor of k^4 is of course also familiar from our discussion of dipole radiation.

Summing and Averaging Polarizations

Let's discuss further what to do with polarization. We set up our polarization vectors according to the picture and table below.

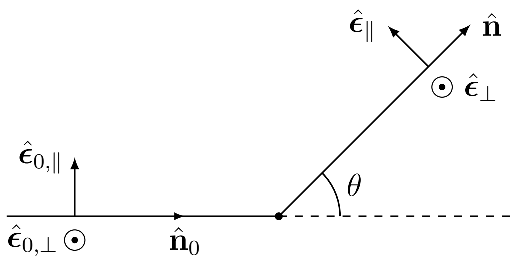


Table of $|\hat{\boldsymbol{\epsilon}}_0 \cdot \hat{\boldsymbol{\epsilon}}|^2$

$\hat{\boldsymbol{\epsilon}}_0$ \\ $\hat{\boldsymbol{\epsilon}}$	$\parallel$	$\perp$
$\parallel$	$\cos^2\theta$	0
$\perp$	0	1

We can either send in unpolarized light or knowingly send in light of a particular polarization. Similarly, we choose to either detect light of a particular polarization or we choose to detect light of all polarizations. Usually, we send in unpolarized light and detect all polarizations. In this case, we should average over initial polarizations and sum over final polarizations.

Assume the initial beam is unpolarized. Then we average over initial polarizations to get

$$\frac{d\sigma}{d\Omega}_\parallel = \frac{1}{2}\left[\frac{d\sigma}{d\Omega}(\parallel\to\parallel) + \frac{d\sigma}{d\Omega}(\perp\to\parallel)\right] = \frac{1}{2}\sigma_0 \cos^2\theta \tag{20}$$

$$\frac{d\sigma}{d\Omega}_\perp = \frac{1}{2}\left[\frac{d\sigma}{d\Omega}(\parallel\to\perp) + \frac{d\sigma}{d\Omega}(\perp\to\perp)\right] = \frac{1}{2}\sigma_0, \tag{21}$$

where, in the dielectric sphere example, we had

$$\sigma_0 = (ka)^4 a^2 \left| \frac{\varepsilon - 1}{\varepsilon + 2} \right|^2. \tag{22}$$

We obtain the total differential cross section by summing over the final polarizations:

$$\frac{d\sigma}{d\Omega}(\text{ave. initial, sum final}) = \left(\frac{1 + \cos^2\theta}{2} \right) \sigma_0. \tag{23}$$

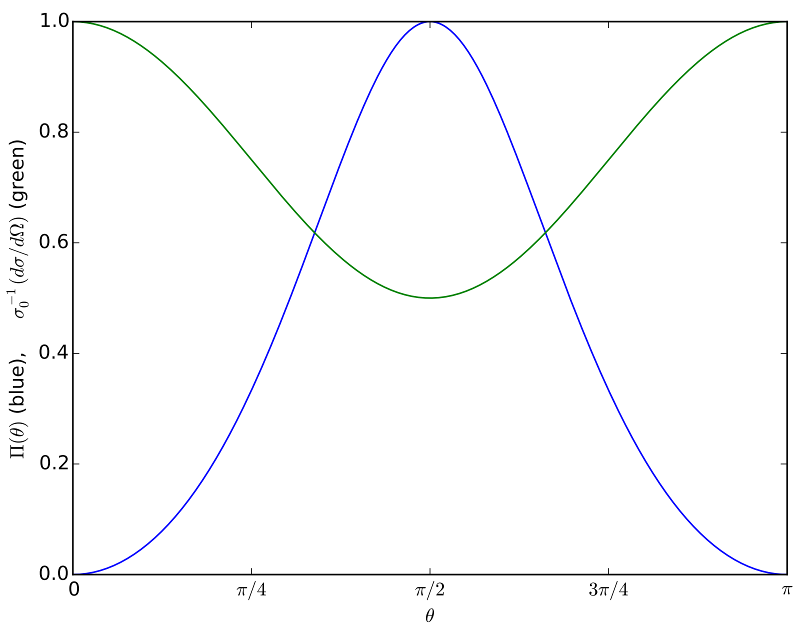
We see there's a slight maximum for forward and backwards scattering. The total cross section is

$$\sigma_{\text{tot}} = \int \frac{d\sigma}{d\Omega}\, d\cos\theta\, d\phi = \frac{2\pi}{2}\sigma_0 \left[2 + \frac{2}{3} \right] = \frac{8\pi}{3}\sigma_0. \tag{24}$$

It's interesting to ask what the difference is between out-of-plane and in-plane scattering cross sections, i.e. is the scattered light dominated by a particular polarization? Our metric for this is

$$\Pi(\theta) \equiv \frac{\frac{d\sigma}{d\Omega}_\perp - \frac{d\sigma}{d\Omega}_\parallel}{\frac{d\sigma}{d\Omega}_\perp + \frac{d\sigma}{d\Omega}_\parallel} = \frac{1 - \cos^2\theta}{1 + \cos^2\theta} = \frac{\sin^2\theta}{1 + \cos^2\theta}. \tag{25}$$

Observe that $\Pi(0) = \Pi(\pi) = 0$ and $\Pi(\pm\pi/2) = 1$, telling us that the forward and backward maxima are unpolarized, but light scattered at $\pi/2$ is completely polarized. This explains why light from the sky scattered perpendicular to the sun's rays is completely polarized!



Perfectly Conducting Sphere

For a perfectly conducting sphere we take the $\varepsilon \to \infty$, $\mu \to 0$ limit of a dielectric, magnetically permeable sphere. The limit of the induced dipole moments is

$$\mathbf{p} = \left(\frac{\varepsilon - 1}{\varepsilon + 2}\right) a^3 E_0 \hat{\boldsymbol{\epsilon}}_0 \to a^3 E_0 \hat{\boldsymbol{\epsilon}}_0. \tag{26}$$

$$\mathbf{m} = \left(\frac{\mu - 1}{\mu + 2}\right) a^3 \mathbf{B}_0 = \left(\frac{\mu - 1}{\mu + 2}\right) a^3 (\hat{\mathbf{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0) E_0 \to -\frac{1}{2} a^3 \hat{\mathbf{n}}_0 \times \mathbf{p}. \tag{27}$$

We'll get radiation through both the electric and the magnetic dipole moment, so the scattered electric field is the sum

$$\mathbf{E}_{\text{out}} = \mathbf{E}_{\text{E1}} + \mathbf{E}_{\text{M1}}.$$

This is going to cause interference!

The electric and magnetic dipole fields are

$$\mathbf{E}_{\text{E1}} = -k^2 \frac{e^{ikr}}{r} \hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \mathbf{p}) = -k^2 a^3 E_0 \frac{e^{ikr}}{r} \hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}}_0). \tag{28}$$

$$\mathbf{E}_{\text{M1}} = -k^2 \frac{e^{ikr}}{r} \hat{\mathbf{n}} \times \mathbf{m} = \frac{k^2 a^3 E_0}{2} \frac{e^{ikr}}{r} \hat{\mathbf{n}} \times (\hat{\mathbf{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0). \tag{29}$$

The differential cross section is

$$\frac{d\sigma}{d\Omega}(\hat{\boldsymbol{\epsilon}}, \hat{\mathbf{n}}, \hat{\boldsymbol{\epsilon}}_0, \hat{\mathbf{n}}_0) = k^4 a^6 \left| \hat{\boldsymbol{\epsilon}}^* \cdot \left(\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}}_0) - \frac{1}{2} \hat{\mathbf{n}} \times (\hat{\mathbf{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0) \right) \right|^2 \tag{30}$$

$$= k^4 a^6 \left| \hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0 - \frac{1}{2} (\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}})^* \cdot (\hat{\mathbf{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0) \right|^2 \tag{31}$$

Recall our diagram of the polarization and normal vector orientations. This gives us the various dot products in the tables below.

Table of $\hat{\boldsymbol{\epsilon}} \cdot \hat{\boldsymbol{\epsilon}}_0$

$\hat{\boldsymbol{\epsilon}}_0$ \ $\hat{\boldsymbol{\epsilon}}$	$\parallel$	$\perp$
$\parallel$	$\cos\theta$	0
$\perp$	0	1

Table of $(\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}}) \cdot (\hat{\mathbf{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0)$

$\hat{\boldsymbol{\epsilon}}_0$ \ $\hat{\boldsymbol{\epsilon}}$	$\parallel$	$\perp$
$\parallel$	1	0
$\perp$	0	$\cos\theta$

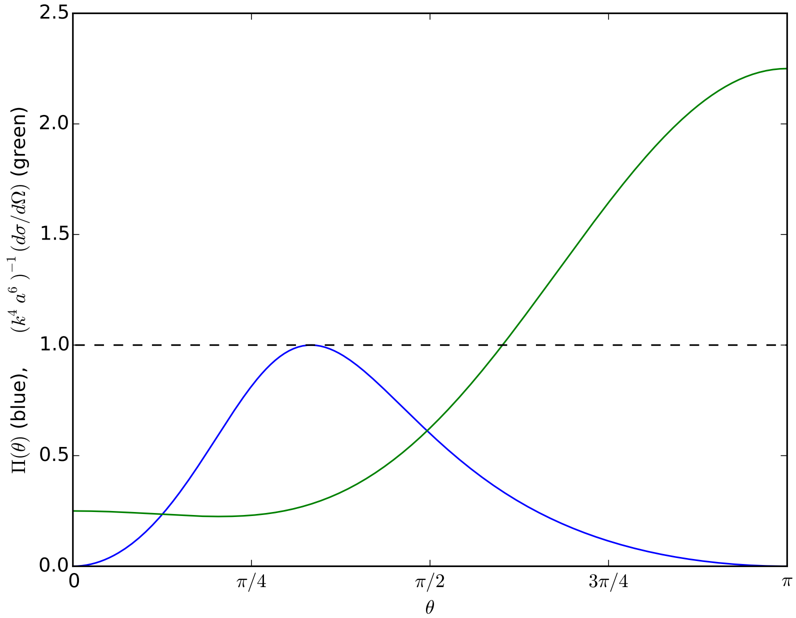
Plugging these back into (31) and averaging over initial polarizations yields

$$\frac{d\sigma}{d\Omega}_{\parallel} = \frac{1}{2} k^4 a^6 \left| \cos\theta - \frac{1}{2} \right|^2 \tag{32}$$

$$\frac{d\sigma}{d\Omega}_{\perp} = \frac{1}{2} k^4 a^6 \left| 1 - \frac{1}{2} \cos\theta \right|^2. \tag{33}$$

$$\frac{d\sigma}{d\Omega}_{\text{tot}} = \frac{1}{8} k^4 a^6 \left(5\cos^2\theta - 8\cos\theta + 5 \right). \tag{34}$$

The differential cross section and $\Pi(\theta)$ are plotted below.



The light is now completely polarized at $\theta = \pi/3$. Moreover, the scattering has a large peak at $\theta = \pi$.

3.2 Scattering off Fluctuations

In the previous section, we were interested in scattering off dielectrics and conductors, and we took the long wavelength limit to make progress. We now consider another practical limit:

$$|\varepsilon - 1|, |\mu - 1| \ll 1, \tag{35}$$

so that what we scatter off of can be considered a fluctuation away from the vacuum. This is really a very general case. If ε and μ are substantial and uniform over distances large compared to a wavelength, then the light doesn't scatter; it reflects and refracts.

Recall that in media without sources the equations of electromagnetism are

$$\mathbf{D} = \varepsilon \mathbf{E} = \mathbf{E} + 4\pi \mathbf{P} \qquad \mathbf{B} = \mu \mathbf{H} = \mathbf{H} + 4\pi \mathbf{M}. \tag{36}$$

$$\nabla \cdot \mathbf{D} = 0 \qquad \nabla \cdot \mathbf{B} = 0 \tag{37}$$

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \qquad \nabla \times \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t}. \tag{38}$$

We can use these to write down a wave equation for $\mathbf{D}$. First we take second derivatives:

$$\frac{1}{c^2}\frac{\partial^2 \mathbf{D}}{\partial t^2} = \frac{1}{c}\frac{\partial}{\partial t}(\boldsymbol{\nabla}\times\mathbf{H}) \tag{39}$$

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times(\mathbf{D}-\mathbf{E})) = -\boldsymbol{\nabla}^2\mathbf{D} + \boldsymbol{\nabla}\times\left(\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}\right). \tag{40}$$

Combining these equations into a d'Alembertian yields

$$\left[\boldsymbol{\nabla}^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right]\mathbf{D} = -\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times(\mathbf{D}-\mathbf{E})) + \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{\nabla}\times(\mathbf{B}-\mathbf{H}). \tag{41}$$

If we're away from the fluctuation (i.e. if $\varepsilon = \mu = 1$), then the right hand side is zero. More generally, if we assume the right hand side is some known source $\mathbf{K}$ with $e^{-i\omega t}$ time dependence, the Fourier transform of (41) is

$$\left(\boldsymbol{\nabla}^2 + k^2\right)\mathbf{D} = \mathbf{K}, \quad k = \frac{\omega}{nc}, \tag{42}$$

where n is the index of refraction. This is the inhomogeneous Helmholtz equation, whose solutions we know are

$$\mathbf{D}(\mathbf{x}) = \mathbf{D}_0(\mathbf{x}) + \int d^3x'\, G(\mathbf{x},\mathbf{x}',k)\mathbf{K}(\mathbf{x}'), \tag{43}$$

$$G(\mathbf{x},\mathbf{x}',k) = -\frac{1}{4\pi}\frac{e^{ikR}}{R} \simeq -\frac{1}{4\pi}\frac{e^{ikr}}{r}e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}, \tag{44}$$

where $\mathbf{D}_0(\mathbf{x})$ is the incident wave, $R = |\mathbf{x}-\mathbf{x}'|$, and we have approximated G in the radiation zone.

For simplicity, let's consider for the moment only the electric part of $\mathbf{K}$. In the radiation zone, the scattered $\mathbf{D}$ field is

$$\mathbf{D}_{\text{sc}} = \frac{1}{4\pi}\frac{e^{ikr}}{r}\int d^3x'\, e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}\boldsymbol{\nabla}'\times(\boldsymbol{\nabla}'\times(\mathbf{D}(\mathbf{x}')-\mathbf{E}(\mathbf{x}'))). \tag{45}$$

We integrate by parts twice to apply $\boldsymbol{\nabla}'$ to $e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}$, with the result

$$\mathbf{D}_{\text{sc}} = \frac{e^{ikr}}{r}\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\mathbf{f}_s(\hat{\mathbf{n}})). \tag{46}$$

$$\mathbf{f}_{\text{sc}}(\hat{\mathbf{n}}) = -\frac{k^2}{4\pi}\int d^3x'\, e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}[\mathbf{D}(\mathbf{x}')-\mathbf{E}(\mathbf{x}')], \tag{47}$$

where $\mathbf{f}_{\text{sc}}(\hat{\mathbf{n}})$ is the scattering amplitude.

Now we go on to calculate the differential cross section. The power distribution is

$$\frac{dP}{d\Omega} = \frac{cr^2}{8\pi}\hat{\mathbf{n}}\cdot[\mathbf{D}^*_{\text{sc}}\times\mathbf{B}_{\text{sc}}] = \frac{cr^2}{8\pi}|\mathbf{D}_{\text{sc}}|^2. \tag{48}$$

For a certain polarization, we replace $|\mathbf{D}_{\text{sc}}|^2$ with $|\hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{D}_{\text{sc}}|^2$. Including the polarization actually simplifies things since

$$\hat{\boldsymbol{\epsilon}}^* \cdot (\hat{\mathbf{n}} \times (\hat{\mathbf{n}} \times \mathbf{f}_s)) = -\hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{f}_s. \tag{49}$$

The scattering cross section in a particular polarization is then

$$\boxed{\frac{d\sigma}{d\Omega} = \frac{|\hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{f}_s(\hat{\mathbf{n}})|^2}{|\mathbf{D}_0|^2}.} \tag{50}$$

But we're in trouble: $\mathbf{f}_s$ depends on $\mathbf{D}(\mathbf{x}')$, and we don't know $\mathbf{D}(\mathbf{x}')$.

What we do is essentially the Born approximation, but in a version due to Rayleigh from 1881. We substitute ε for

$$\delta\varepsilon(\mathbf{x}) = \varepsilon(\mathbf{x}) - 1. \tag{51}$$

Assuming the incident wave is a plane wave

$$\mathbf{E} = E_0 \hat{\boldsymbol{\epsilon}}_0 e^{i\mathbf{k}_0 \cdot \mathbf{x}}, \tag{52}$$

then

$$\mathbf{D} = \varepsilon(\mathbf{x}) E_0 \hat{\boldsymbol{\epsilon}}_0 e^{i\mathbf{k}_0 \cdot \mathbf{x}} = \mathbf{E} + \delta\varepsilon(\mathbf{x}) E_0 \hat{\boldsymbol{\epsilon}}_0 e^{i\mathbf{k}_0 \cdot \mathbf{x}}. \tag{53}$$

Plugging this into the scattering amplitude yields

$$\mathbf{f}_{\text{sc}} = -\frac{k^2 E_0}{4\pi} \hat{\boldsymbol{\epsilon}}_0 \int d^3x' \, e^{i(\mathbf{k}_0 - \mathbf{k}) \cdot \mathbf{x}'} \delta\varepsilon(\mathbf{x}'). = -\frac{k^2 E_0}{4\pi} \hat{\boldsymbol{\epsilon}}_0 \delta\varepsilon(\mathbf{q}), \tag{54}$$

where $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}$ is the wave number transfer in the scattering. Finally plugging this into the differential cross section yields

$$\frac{d\sigma}{d\Omega}(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\epsilon}}_0) = \frac{k^4}{16\pi^2} |\delta\varepsilon(\mathbf{q}) \hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \tag{55}$$

Throughout the derivation we were considering the electric part of the scattering only. The treatment of the magnetic part is similar, with the result that the total differential cross section is modified to

$$\boxed{\frac{d\sigma}{d\Omega}(\hat{\boldsymbol{\epsilon}}, \hat{\boldsymbol{\epsilon}}_0) = \frac{k^4}{16\pi^2} |\delta\varepsilon(\mathbf{q}) \hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0 + \delta\mu(\mathbf{q}) (\hat{\mathbf{n}} \times \hat{\boldsymbol{\epsilon}}^*) \cdot (\hat{\mathbf{n}}_0 \times \hat{\boldsymbol{\epsilon}}_0)|^2.} \tag{56}$$

In summary, the physical picture is the following. In some limited region of space we have fluctuations in the permittivity and permeability $0 \neq \delta\varepsilon, \delta\mu \ll 1$. Crucially, the waves scatter off the fluctuations; it is $\delta\varepsilon$ and $\delta\mu$ that appear in the differential cross section. In general, the $\mathbf{q}$ dependence of $\delta\varepsilon$ and $\delta\mu$ will give more complicated angular dependence than the spheres we considered in the previous section.

Example 4. If $\delta\varepsilon = 0$ everywhere outside a region of length scale d and $qd \ll 1$, then

$$\delta\varepsilon(\mathbf{q}) = \int d^3x'\, e^{i\mathbf{q}\cdot\mathbf{x}'}\delta\varepsilon(\mathbf{x}') \simeq \int d^3x'\, \delta\varepsilon(\mathbf{x}'). \tag{57}$$

If the scatterer is a uniform sphere of radius a, then the differential cross section is

$$\frac{d\sigma}{d\Omega} = \frac{k^4}{16\pi^2}\left(\frac{4\pi a^3}{3}(\varepsilon - 1)\right)^2 |\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 = k^4 a^6 \left(\frac{\varepsilon - 1}{3}\right)^2 |\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \tag{58}$$

This is equal to the exact differential cross section from the previous section

$$\frac{d\sigma}{d\Omega} = k^4 a^6 \left(\frac{\varepsilon - 1}{\varepsilon + 2}\right)^2 |\hat{\boldsymbol{\epsilon}}^* \cdot \hat{\boldsymbol{\epsilon}}_0|^2 \tag{59}$$

in the limit $\varepsilon \to 1$.

Many Scatterers

Suppose the scatterer is at some $\mathbf{x}_0$, not necesarily at the origin. Then

$$\mathbf{E}_{\text{in}} = \hat{\boldsymbol{\epsilon}}_0 E_0 e^{ik\hat{\mathbf{n}}_0\cdot\mathbf{x}_0}, \tag{60}$$

where $\hat{\mathbf{n}}_0$ is the beam direction of $\mathbf{E}$ at the scatterer. This will induce a response in the dielectric with the same phase factor

$$\mathbf{J}(\mathbf{x}') \propto e^{ik\hat{\mathbf{n}}_0\cdot\mathbf{x}_0}. \tag{61}$$

If we define $\mathbf{x}'' = \mathbf{x}' - \mathbf{x}_0$, then

$$\mathbf{J}(\mathbf{x}') = e^{ik\hat{\mathbf{n}}_0\cdot\mathbf{x}_0}\mathbf{J}(\mathbf{x}'') \tag{62}$$

$$|\mathbf{x} - \mathbf{x}'| = |\mathbf{x} - \mathbf{x}_0 - \mathbf{x}''| \simeq r - \hat{\mathbf{n}} \cdot (\mathbf{x}_0 + \mathbf{x}''). \tag{63}$$

With these we can rewrite the vector potential induced by the scatterer as

$$\mathbf{A}(\mathbf{x}) = \frac{1}{c}\int \mathbf{J}(\mathbf{x}')\frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|}\, d^3x' \tag{64}$$

$$\simeq \frac{e^{ikr}}{cr}\int \mathbf{J}(\mathbf{x}'')e^{ik\hat{\mathbf{n}}_0\cdot\mathbf{x}_0}e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}_0}e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}''}\, d^3x'' \tag{65}$$

$$= \frac{e^{ikr}}{cr}e^{i\mathbf{q}\cdot\mathbf{x}_0}\int \mathbf{J}(\mathbf{x}'')e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}''}\, d^3x''\,, \tag{66}$$

where $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}$ is the wave number transferred. This is the same as what the vector potential would be if the scatterer were at the origin, except that we've picked up the phase $e^{i\mathbf{q}\cdot\mathbf{x}_0}$. It follows that the multipole moments effectively pick up the same phase factor.

If we have N scatterers then the scattering cross section is

$$\frac{d\sigma}{d\Omega} = \frac{k^4}{|E_0|^2} \left| \sum_j \hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{p}_j e^{i\mathbf{q}\cdot\mathbf{x}_j} \right|^2. \tag{67}$$

The scattering amplitude for each scatterer is just

$$\mathbf{f}(\hat{\mathbf{n}}, \mathbf{x}_j) = e^{i\mathbf{q}\cdot\mathbf{x}_j} \mathbf{f}(\hat{\mathbf{n}}, 0). \tag{68}$$

To simplify things, we assume that all the scatterers are identical, so $\mathbf{p}_i = \mathbf{p}_j$ for all i, j. Then

$$\frac{d\sigma}{d\Omega} = |\hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{f}(\hat{\mathbf{n}}, 0)|^2 \left| \sum_j e^{i\mathbf{q}\cdot\mathbf{x}_j} \right|^2 = \frac{d\sigma_0}{d\Omega}(\hat{\boldsymbol{\epsilon}}) F(\mathbf{q}), \tag{69}$$

where $d\sigma_0/d\Omega$ is the differential cross section for a single scatterer at the origin and

$$F(\mathbf{q}) \equiv \left| \sum_j e^{i\mathbf{q}\cdot\mathbf{x}_j} \right|^2 \tag{70}$$

is called the ***form factor*** or ***structure factor***. The form factor contains all the information about the spatial distribution of the dipoles.

Example 5. If we have a regular array of scatterers, then the mathematics is exactly the same as the line of dipoles we analyzed on first problem set. The form factor can be shown to be

$$F(\mathbf{q}) = N^2 \left[\frac{\sin^2\left(\frac{N_1 q_1 a}{2}\right)}{N_1^2 \sin^2\left(\frac{q_1 a}{2}\right)} \frac{\sin^2\left(\frac{N_2 q_2 a}{2}\right)}{N_2^2 \sin^2\left(\frac{q_2 a}{2}\right)} \frac{\sin^2\left(\frac{N_3 q_3 a}{2}\right)}{N_3^2 \sin^2\left(\frac{q_3 a}{2}\right)} \right], \tag{71}$$

where a is the lattice spacing, and N_i is the number of lattice sites along the ith axis. This gives rise to Bragg peaks.

Example 6. What if we have a random array of scatterers? In general, we can explicitly do the squaring in the form factor to get

$$F(\mathbf{q}) = \sum_{ij} e^{i\mathbf{q}\cdot(\mathbf{x}_i - \mathbf{x}_j)}. \tag{72}$$

This is a sum over phase factors weighted by $\mathbf{x}_i - \mathbf{x}_j$. But if the values of $\mathbf{x}_i - \mathbf{x}_j$ are random, then we'll get totally destructive interference:

$$\sum_{i \neq j} e^{i\mathbf{q}\cdot(\mathbf{x}_i - \mathbf{x}_j)} = 0. \tag{73}$$

The remaining sum over the $i = j$ terms is just

$$F(\mathbf{q}) = \sum_{i=1}^{N} 1 = N. \tag{74}$$

The differential cross section is then N times the differential cross section of a single scatterer. Makes sense. This is called ***incoherent scattering***.

However, this is not the whole story. If $\mathbf{q} = 0$, i.e. if the direction of scattering is in the same direction as the beam, then

$$F(\mathbf{q}) = \left|\sum_{i=1}^{N} 1\right|^2 = N^2. \tag{75}$$

The scattering amplitude is enhanced by N^2 rather than N! This is called ***coherent scattering***. It is why you can see an object through air!

More precisely, if d is the size of the scattering region, then we need $qd \ll 1$ to make all the exponentials $e^{i\mathbf{q}\cdot(\mathbf{x}_i - \mathbf{x}_j)}$ close to one and get coherent scattering. Since $|q| = 2k\sin(\theta/2)$, the condition $qd \ll 1$ is equivalent to $k\theta d \ll 1$, or

$$\theta \ll \frac{1}{kd} = \frac{\lambda}{2\pi d}. \tag{76}$$

If, for example, we scatter visible light on $1\,\mathrm{m}^3$ of air, then

$$\theta \ll \frac{500\,\mathrm{nm}}{2\pi \cdot 1m} \simeq 10^{-7}\,\mathrm{rad}, \tag{77}$$

which is a tiny window!

We now consider the power lost in a beam due to scattering. Imagine looking at a slab of scatterers of area A and depth Δx, with particle density $n = N/A\Delta x$. Then

$$\Delta P_{\text{loss}} = -\left[\text{incident flux} = \frac{\text{power in}}{\text{unit area}}\right] \times \left[\sigma = \frac{\text{power loss per molecule}}{\text{incident flux}}\right] \times N \tag{78}$$

$$\Delta P = -\frac{P}{A}\sigma n A \Delta x \tag{79}$$

$$\implies \quad \frac{\Delta P}{\Delta x} = -n\sigma P, \tag{80}$$

where n is the number density. This gives exponential decay of the power!

$$P(x) = P_0 e^{-n\sigma x} = P_0 e^{-x/\lambda}, \tag{81}$$

where $\lambda \equiv 1/n\sigma$ is the ***attenuation length***. Recall Rayleigh's law $\sigma \propto k^4$. Clearly, the attenuation length is much smaller for high frequency light, so blue light scatters away more. Hence the sun looks redder at sunset, when its rays travel through more of the atmosphere to reach you.

We can use this to measure the density of air, for example. The incident electric field induces dipole moments $\mathbf{p} = \gamma_{\text{mol}}\mathbf{E}$ in each molecule, where γ_{mol} is the ***molecular polarizability***. From our discussion of dipole scattering, we know the cross section per molecule for unpolarized light is

$$\sigma = \frac{8\pi}{3}\frac{k^4|\mathbf{p}|^2}{|\mathbf{E}|^2} = \frac{8\pi}{3}k^4\gamma_{\text{mol}}^2. \tag{82}$$

To relate $\gamma_{\rm mol}^2$ to something more familiar, consider the total polarization of the material $\mathbf{P} = n\mathbf{p}$. We know

$$\varepsilon \mathbf{E} = \mathbf{D} = \mathbf{E} + 4\pi \mathbf{P} = (1 + 4\pi n \gamma_{\rm mol})\mathbf{E}, \tag{83}$$

so $\gamma_{\rm mol} = (\varepsilon - 1)/4\pi n$ and

$$\lambda = \frac{1}{n\sigma} = \frac{6\pi}{k^4}\frac{n}{(\varepsilon-1)^2} \tag{84}$$

To go a step further, we can write $\varepsilon = n_{\rm ref}^2$, where $n_{\rm ref}$ is the index of refraction (not to be confused with the number density n). The index of refraction of air differs from 1 by about one part in a thousand. Given that you can see a hundred miles on a clear day, we can estimate

$$n_{\rm air} = (100\,{\rm mi})\frac{(2\pi)^4}{6\pi(500\,{\rm nm})^4}(n_{\rm ref}-1)^2(n_{\rm ref}+1)^2 = 9\times 10^{26}\,{\rm m}^{-3}, \tag{85}$$

which, given the true value of $0.25 \times 10^{26}\,{\rm m}^{-3}$, is pretty good for such a rough estimate!

Continuous Distributions

Imagine replacing the discrete distribution of scatterers by a continuous distribution:

$$F(\mathbf{q}) = \left|\sum_j e^{i\mathbf{q}\cdot\mathbf{x}_j}\right|^2 \to \left|\int d^3x\, \delta n(x) e^{i\mathbf{q}\cdot\mathbf{x}}\right|^2,$$

where δn is the fluctuation in local matter density

$$\delta n(x) = n(x) - \bar{n}, \qquad \bar{n} = \frac{1}{V}\int d^3x\, n(x).$$

To go a step further, let us average the form factor over a statistical ensemble. Then

$$\begin{aligned} F(\mathbf{q}) &= \left\langle \left|\int d^3x\, \delta n(x) e^{i\mathbf{q}\cdot\mathbf{x}}\right|^2 \right\rangle \\ &= \left\langle \int d^3x\, \delta n(x) e^{i\mathbf{q}\cdot\mathbf{x}} \int d^3y\, \delta n(y) e^{-i\mathbf{q}\cdot\mathbf{y}} \right\rangle \\ &= \int d^3x\, d^3y\, e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})} \left\langle \delta n(\mathbf{x}) \delta n(\mathbf{y}) \right\rangle \end{aligned}$$

In the $\mathbf{q} \to 0$ limit, we have

$$F(\mathbf{q}\to 0) = \int d^3x \int d^3y \left\langle (n(x)-\bar{n})(n(y)-\bar{n}) \right\rangle = \left\langle N^2 \right\rangle - \left\langle N \right\rangle^2.$$

For an ideal gas this is proportional to N. The phenomenon of coherent scattering has been lost in the transition to a continuous matter density. For a more general system than an ideal gas, however, we have

$$F(\mathbf{q}\to 0) = \left\langle N^2\right\rangle - \left\langle N\right\rangle^2 = \overline{(\Delta N)^2} = \frac{N^2 kT}{V}\kappa_T,$$

where κ_T is the isothermal compressibility. This diverges at the critical point in a phase diagram, and so we get a large amount of scattering. This phenomenon is called ***critical opalescence***.

If the system is translationally invariant, then $\langle \delta n(\mathbf{x})\delta n(\mathbf{y})\rangle = \langle \delta n(\mathbf{x}-\mathbf{y})\delta n(0)\rangle$. Then the form factor reduces to

$$F(\mathbf{q}) = \int d^3(x-y)\, d^3y\, e^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{y})}\left\langle \delta n(\mathbf{x}-\mathbf{y}), \delta n(0)\right\rangle = V\int d^3r\, e^{i\mathbf{q}\cdot\mathbf{r}}\left\langle \delta n(\mathbf{r})\delta n(0)\right\rangle.$$

Thus, the form factor is the Fourier transform of the correlation function $\langle \delta n(r)\delta n(0)\rangle$. The key point is that we can learn about the structure of matter by scattering off it! We measure the form factor, and deduce the correlation function.

The theme of this section is that light scatters off fluctuations. Here we've considered density fluctuations, but you can consider other kinds of fluctuations. For example, one can scatter off spins

$$F_{\text{spin}}(\mathbf{q}) = \int d^3r\, \left\langle (\sigma(r)-\overline{\sigma})(\sigma(0)-\overline{\sigma})\right\rangle e^{i\mathbf{q}\cdot\mathbf{r}}.$$

The $\mathbf{q}\to 0$ limit yields

$$F_{\text{spin}} = \left\langle \sigma^2\right\rangle - \left\langle \sigma\right\rangle^2 = \chi = \left.\frac{\partial M}{\partial n}\right|_{n=0},$$

where χ is the ***magnetic susceptibility***. Again, we find that we can learn about materials by scattering off them. The magnetization decreases with T, eventually reaching zero at some T_c. At T_c the magnetic susceptibility χ diverges, so again we get critical opalescence.

4 Diffraction

Diffraction is scattering on a hole. We send in a wave, it passes through a hole in a screen and scatters, and we're interested in the radiation pattern on the other side. As we'll see later, the analogy is made even more concrete by Babinet's principle.

4.1 Boundary Value Problem for Diffraction

We'll begin our discussion of scattering with scalar fields for simplicity. Consider a harmonically oscillating scalar field

$$\psi(x,t) = \psi(x)e^{-i\omega t}, \tag{1}$$

and let it obey the vacuum Helmholtz equation

$$\big(\boldsymbol{\nabla}^2 + k^2\big)\psi = 0. \tag{2}$$

The key difference between our previous discussions and the present situation is that now we want to consider some nontrivial boundary conditions. We'll need a Green's function

$$\big(\boldsymbol{\nabla}^2 + k^2\big)G(\mathbf{x}, \mathbf{x}', k) = -\delta^3(\mathbf{x} - \mathbf{x}'), \tag{3}$$

and we'll need to recall Green's theorem:

$$\int_V \big[G(\mathbf{x}, \mathbf{x}')\boldsymbol{\nabla}'^2\psi(\mathbf{x}') - \psi(\mathbf{x}')\boldsymbol{\nabla}'^2 G(\mathbf{x}, \mathbf{x}')\big]\, d^3x' = \int_S \left[G(\mathbf{x}, \mathbf{x}')\frac{\partial \psi}{\partial n'} - \psi\frac{\partial G}{\partial n'}(\mathbf{x}, \mathbf{x}')\right] da', \tag{4}$$

where $S = \partial V$ and $\partial/\partial n'$ is an outward normal derivative. Applying the Helmholtz equation to $\boldsymbol{\nabla}'^2\psi$ and $\boldsymbol{\nabla}'^2 G$ yields

$$\psi(\mathbf{x}) = \int_S da'\, \hat{\mathbf{n}}' \cdot [\psi\boldsymbol{\nabla}' G - G\boldsymbol{\nabla}'\psi], \tag{5}$$

where $\hat{\mathbf{n}}'$ is an *inward* directed normal vector. We've switched the orientation of $\hat{\mathbf{n}}'$ for future convenience.

This is, of course, not yet a solution for ψ. We would have to specify both ψ and $\boldsymbol{\nabla}'\psi$ on the boundary S in order for it to be so. These are called Cauchy boundary conditions, and are mathematically inconsistent.

Rather, to keep going we remember our Green's function

$$G(\mathbf{x}, \mathbf{x}') = \frac{1}{4\pi}\frac{e^{ikR}}{R}, \tag{6}$$

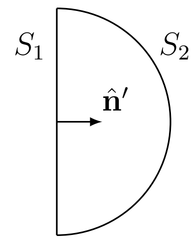
where $\mathbf{R} = \mathbf{x} - \mathbf{x}'$. Taking the gradient yields

$$\boldsymbol{\nabla}'\left(\frac{e^{ikR}}{R}\right) = \left(ik - \frac{1}{R}\right)\frac{e^{ikR}}{R}\boldsymbol{\nabla}' R = -\left(ik - \frac{1}{R}\right)\frac{e^{ikR}}{R}\frac{\mathbf{R}}{R}. \tag{7}$$

Plugging this back into the equation for ψ yields

$$\psi(\mathbf{x}) = -\frac{1}{4\pi}\int_S da'\, \hat{\mathbf{n}}' \cdot \left[\boldsymbol{\nabla}'\psi + \psi(\mathbf{x}')\left(ik - \frac{1}{R}\right)\frac{\mathbf{R}}{R}\right]\frac{e^{ikR}}{R}. \tag{8}$$

We now specialize to the following shape for S. We choose it to be an infinite plane screen S_1 closed off by a hemisphere S_2 of infinite radius. We live inside this hemisphere and the normal vector $\hat{\mathbf{n}}'$ points inside as well.



We'll kill off the S_2 integrals and set $\psi = 0$ everywhere on S_1 except for on the hole in the screen. On S_2, we have large r' so ψ is approximately an outgoing spherical wave

$$\psi \sim \frac{e^{ikr'}}{r'} \quad \Longrightarrow \quad \boldsymbol{\nabla}'\psi \sim \left(ik - \frac{1}{r'}\right)\frac{e^{ikr'}}{r'}\frac{\mathbf{x}'}{r'} \tag{9}$$

and also $\mathbf{R} \sim -\mathbf{x}'$. Thus, the S_2 integral is

$$\int_{S_2}\left[\left(ik - \frac{1}{r'}\right) - \left(ik - \frac{1}{r'}\right)\right]\left(\frac{e^{ikr'}}{r'}\right)^2 \hat{\mathbf{n}}' \cdot \frac{\mathbf{x}'}{r'}\, da' = 0. \tag{10}$$

We're left with the integral over S_1, which in the radiation zone $kR \gg 1$ is

$$\begin{aligned} \psi(\mathbf{x}) &= \int_{S_1} da'\, \hat{\mathbf{n}}' \cdot [\psi \boldsymbol{\nabla}' G - G \boldsymbol{\nabla}'\psi] && (11) \\ &= -\frac{1}{4\pi}\int_{S_1} da'\, \hat{\mathbf{n}}' \cdot \left[\boldsymbol{\nabla}'\psi + ik\psi(\mathbf{x}')\frac{\mathbf{R}}{R}\right]\frac{e^{ikR}}{R}. && (12) \end{aligned}$$

This is the ***Kirchhoff integral formula***.

To summarize, the game we're playing is to take the screen, specify information on the screen, and calculate radiation far from the screen. The issues are the following: we need mathematical consistency (not Cauchy boundary conditions), we need to do practical calculations, and we will need to know how to go from scalar fields to vector fields.

We'll address these issues in turn. For consistency, we need to choose Dirichlet or Neumann boundary conditions. For Dirichlet boundary conditions we choose

$$G = G_D, \text{ where } G_D = 0 \text{ on } S_1. \tag{13}$$

For Neumann boundary conditions we choose

$$G = G_N, \text{ where } \frac{\partial G_N}{\partial n'} = 0 \text{ on } S_1. \tag{14}$$

Let's focus on Dirichlet boundary conditions first. We can use the method of images to find G_D! The image charge goes at the mirror image $\mathbf{x}''$ of $\mathbf{x}'$ behind the screen:

$$\begin{aligned} \mathbf{x}' &= x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}} + z'\hat{\mathbf{z}} && (15) \\ \mathbf{x}'' &= x'\hat{\mathbf{x}} + y'\hat{\mathbf{y}} - z'\hat{\mathbf{z}}. && (16) \end{aligned}$$

Then the Green's function is

$$G_D = \frac{1}{4\pi}\left[\frac{e^{ikR}}{R} - \frac{e^{ikR''}}{R''}\right], \tag{17}$$

where $R'' = |\mathbf{x} - \mathbf{x}''|$. On the screen S_1, we have $z' = 0$, so $R = R''$ and the Green's function is zero. Furthermore, $(\boldsymbol{\nabla}'^2 + k^2)\big(e^{ikR''}/R''\big) = -4\pi\delta(\mathbf{x} - \mathbf{x}'') = 0$ everywhere past the screen, as it must. Plugging this into the equation for ψ, we obtain

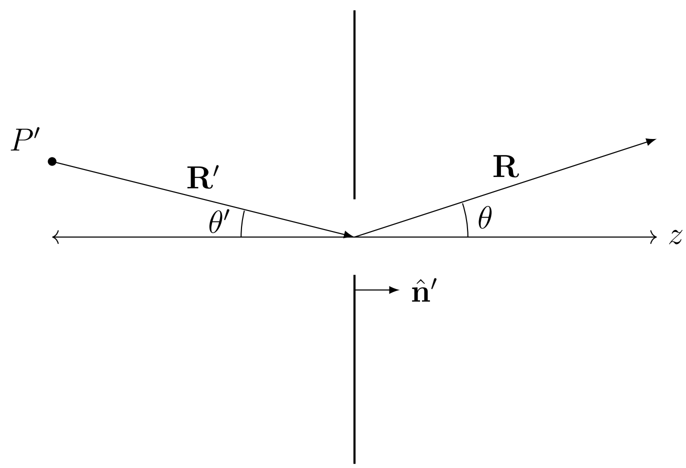
$$\psi_D(\mathbf{x}) = -\frac{ik}{2\pi}\int_{S_1} da'\, \psi(\mathbf{x}')\frac{\hat{\mathbf{n}}' \cdot \mathbf{R}}{R}\frac{e^{ikR}}{R}. \tag{18}$$

The Green's function for Neumann boundary conditions is the same as the Dirichlet Green's function except we replace the minus sign with a plus sign. This results in

$$\psi_N = -\frac{1}{2\pi}\int_{S_1} da'\,\frac{e^{ikR}}{R}\hat{\mathbf{n}}'\cdot\boldsymbol{\nabla}'\psi(\mathbf{x}'). \tag{19}$$

The Kirchhoff integral is the average of these two equations. In all cases, we may assume that we only need to integrate over the aperture in the screen.

These integral formulae can be summarized in the following way. Suppose we have a point source P' behind the screen that emits a spherical wave $\psi_0(\mathbf{R}') = e^{ikR'}/R'$, where R' is a distance measured from P'. Define the angles θ and θ' according to the diagram below.



Each of the integral formulae can be rewritten in the form

$$\psi(\mathbf{x}) = -\frac{ik}{2\pi}\int_{S_1}\frac{e^{ikR}}{R}\frac{e^{ikR'}}{R'}\mathscr{O}(\theta,\theta')\,da'\,. \tag{20}$$

For Dirichlet boundary conditions $\mathscr{O} = \cos\theta$, for Neumann boundary conditions we have $\mathscr{O} = \cos\theta'$, and in the Kirchhoff formula we have $\mathscr{O} = (\cos\theta + \cos\theta')/2$. Even more choices can be made for $\mathscr{O}$; for example the Huygens formula has $\mathscr{O} = 1$. In most practical cases, these end up being the same. We'll largely consider the source and the diffraction pattern to be far from the screen and mostly in line with the aperture, so that $\theta' \simeq \theta \simeq 0$ and $\cos\theta \simeq \cos\theta' \simeq 1$.

4.2 Fraunhofer and Fresnel Diffraction

Let's take the Dirichlet wave function

$$\psi(\mathbf{x}) = \frac{k\cos\theta}{2\pi i}\int_{\text{aperture}} da'\,\psi_0(\mathbf{x}')\frac{e^{ikR}}{R}, \tag{21}$$

In the far field, we expand kR as

$$kR = kr - k\hat{\mathbf{n}} \cdot \mathbf{x}' + \frac{k}{2r}\left[r'^2 - (\hat{\mathbf{n}} \cdot \mathbf{x}')^2\right] + \cdots \tag{22}$$

The first term has size kr, the second term has size kd, the last term has size $kd \cdot (d/r)$. Which terms we keep depends on their relative sizes, of course. There are two common cases that arise from the size of d/r.

In ***Fraunhofer diffraction*** $d/r \ll 1$, so we can throw away the third term in the above expansion. Then

$$\frac{e^{ikR}}{R} = \frac{e^{ikr}}{r} e^{-ik\hat{\mathbf{n}} \cdot \mathbf{x}'}. \tag{23}$$

To get $\psi(\mathbf{x})$ we need some formula for the source $\psi_0(\mathbf{x}')$. If we send in a plane wave $\psi_0(\mathbf{x}') = \psi_0 e^{i\mathbf{k}_0 \cdot \mathbf{x}}$ at some wavevector $\mathbf{k}_0$, then

$$\psi_D = \frac{\psi_0 k}{2\pi i} \frac{e^{ikr}}{r} \cos\theta \int_{\text{aperture}} da' \, e^{i\mathbf{q} \cdot \mathbf{x}'}, \tag{24}$$

where $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}$ as usual.

Example 7. Suppose we have normally incident light on a square hole of side length L. Then $\mathbf{k}_0 = (0, 0, k)$, $\mathbf{x}' = (x', y', 0)$, and $\mathbf{k} = (k_x, k_y, k_z)$, so $\mathbf{q} \cdot \mathbf{x}' = -k_x x' - k_y y'$. Then

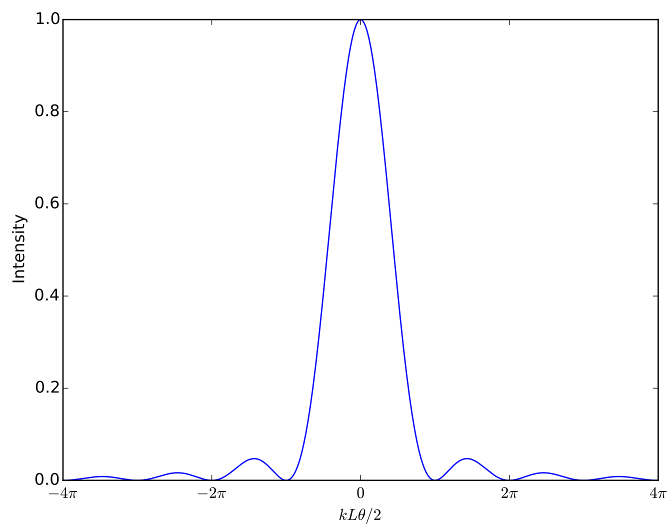
$$\int_{\text{aperture}} da' \; e^{i\mathbf{q} \cdot \mathbf{x}'} = \int_{-L/2}^{L/2} dy' \, e^{-ik_y y'} \int_{-L/2}^{L/2} dx' \, e^{-ik_x x'} \tag{25}$$

$$= L^2 \left[\frac{\sin(k_y L/2)}{k_y L/2}\right] \left[\frac{\sin(k_x L/2)}{k_x L/2}\right] \tag{26}$$

If we look along the $y = 0$ plane, then $k_y = 0$ and $k_x = k \sin\theta \sim k\theta$, so the intensity goes as

$$I \propto \frac{\sin^2(kL\theta/2)}{(kL\theta/2)^2}, \tag{27}$$

shown below.



There are nodes at $kL\theta/2 = n\pi$. The intensity envelope falls as $1/(kL\theta)^2$, so nearly all the intensity falls within the first node.

Note that although we sent in a wave with wavevector $\mathbf{k}_0$, we get wave motion in directions $\mathbf{k} \neq \mathbf{k}_0$ after the wave passes through the aperture. We can think about this via the Heisenberg uncertainty principle. Roughly speaking, $\Delta k_x = k_x = k\theta$ and $\Delta x = L$, so $\Delta k_x \Delta x = k\theta L \sim 2\pi$.

The other situation we might be concerned with is called ***Fresnel diffraction***. We assume $kd^2/r > 1$, so the second term in the expansion of R matters. Let's do an example.

Example 8. Let's suppose we have a screen in the half plane $x < 0$, $z = 0$. We will have to integrate in x' and y' all the way up to infinity. Looking at a point with $y = 0$, we have

$$\mathbf{R} = (x - x')\hat{\mathbf{x}} - y'\hat{\mathbf{y}} + z\hat{\mathbf{z}}, \qquad R = \left[(x-x')^2 + y'^2 + z^2\right]^{1/2} \tag{28}$$

We'll assume z is large, everything else is small, and then integrate the small things from zero to infinity. That's physics for you. In this approximation,

$$\hat{\mathbf{n}}' \cdot \mathbf{R} \simeq \hat{\mathbf{n}}' \cdot \mathbf{z}, \qquad R \simeq z + \frac{1}{2z}\left[(x-x')^2 + y'^2\right]. \tag{29}$$

The amplitude at normal incidence is

$$\psi = \frac{k}{2\pi i}\frac{e^{ikz}}{z} I_0^{1/2} \int_0^\infty dx' \int_{-\infty}^\infty dy' \exp\left(ik\left[\frac{(x-x')^2}{2z} + \frac{y'^2}{2z}\right]\right) \tag{30}$$

We can do the integral over y'; it's just a Gaussian. The integral over x' requires Fresnel integrals.

4.3 Vector Diffraction

Diffraction of vector fields is somewhat subtle; we'll only discuss it very briefly here. The formula for vector scattering is called the ***Vector Smythe-Kirchhoff Formula***. Smythe discovered this formula in the 1940s, so this is extremely recent compared to the rest of classical electrodynamics! The formula is

$$\mathbf{E}(\mathbf{x}) = \frac{1}{2\pi} \boldsymbol{\nabla} \times \int_{\text{aperture}} da' \, \frac{e^{ikR}}{R} \hat{\mathbf{n}}' \times \mathbf{E}(\mathbf{x}') \tag{31}$$

We won't derive this formula, but let's explain the difficulties in doing so. The radiation doesn't just come from the hole, it also comes from the wall. Suppose you have some incident electric and magnetic field $\mathbf{E}_i$ and $\mathbf{B}_i$. If there's no hole, then the electric and magnetic fields on the other side are zero. This means that the screen had to generate an equal and opposite electric and magnetic field to cancel the incoming one: $\mathbf{E} = 0 = \mathbf{E}_i + \mathbf{E}_{\text{sc}}$. Thus, these formulas only apply when our screen is conducting and can generate such fields.

Once we have an aperture, we can argue that the dominant contribution to the electric field is from the hole, but this is not the case for the magnetic field! For the magnetic field we must still consider the screen. This is an essential difference between vector and scalar diffraction and it causes some difficulties.

Diffraction by a Circular Aperture

We will use the Smythe-Kirchhoff formula

$$\mathbf{E}(\mathbf{x}) = \frac{1}{2\pi}\boldsymbol{\nabla} \times \int_{\text{aperture}} da' \, \frac{e^{ikR}}{R}\hat{\mathbf{z}} \times \mathbf{E}_0(\mathbf{x}'). \tag{32}$$

In the Fraunhofer limit,

$$\frac{e^{ikR}}{R} \simeq \frac{e^{ikr}}{r} e^{-ik\hat{\mathbf{n}}\cdot\mathbf{x}'}. \tag{33}$$

and

$$\boldsymbol{\nabla}\left(\frac{e^{ikR}}{R}\right) \simeq \frac{e^{ikR}}{R} ik\hat{\mathbf{n}}, \tag{34}$$

where $\hat{\mathbf{n}}$ points from the origin to the observation point. If we take the incoming wave to be a plane wave $\mathbf{E} = \hat{\boldsymbol{\epsilon}}_0 E_0 e^{i\mathbf{k}_0\cdot\mathbf{x}}$, we have

$$\mathbf{E}(\mathbf{x}) = \frac{ikE_0}{2\pi}\frac{e^{ikr}}{r}\hat{\mathbf{n}} \times (\hat{\mathbf{z}} \times \hat{\boldsymbol{\epsilon}}_0) \int_{\text{aperture}} da' \, e^{i\mathbf{q}\cdot\mathbf{x}'}, \tag{35}$$

where $\mathbf{q} = \mathbf{k}_0 - \mathbf{k}$ and $\mathbf{k} = k\hat{\mathbf{n}}$ as usual. For simplicity, let's assume we the plane wave is normally incident $\mathbf{k}_0 = k\hat{\mathbf{z}}$ and choose coordinates so that the polarization is $\hat{\boldsymbol{\epsilon}}_0 = \hat{\mathbf{x}}$. Then $\hat{\mathbf{z}} \times \hat{\boldsymbol{\epsilon}}_0 = \hat{\mathbf{y}}$. The aperture is in the xy-plane, so $\mathbf{k}_0 \cdot \mathbf{x}' = 0$. All the interesting behavior is in the shape of the aperture, in the integral

$$I = \int_{\text{aperture}} e^{-i\mathbf{k}\cdot\mathbf{x}'} \, da' \, . \tag{36}$$

We take ϕ and θ to be the usual spherical coordinates of $\hat{\mathbf{n}}$ and ϕ' and ρ' to be polar coordinates of $\mathbf{x}'$ in the xy-plane. Explicitly,

$$\hat{\mathbf{n}} = (\hat{\mathbf{x}}\cos\phi + \hat{\mathbf{y}}\sin\phi)\sin\theta + \hat{\mathbf{z}}\cos\theta \tag{37}$$

$$\hat{\mathbf{x}}' = \rho'(\hat{\mathbf{x}}\cos\phi' + \hat{\mathbf{y}}\sin\phi'). \tag{38}$$

Then

$$\hat{\mathbf{n}} \cdot \mathbf{x}' = \rho' \sin\theta\cos\phi\cos\phi' + \sin\theta\sin\phi\sin\phi' = \rho'\sin\theta\cos(\phi - \phi'). \tag{39}$$

Plugging this all in, the integral is

$$I = \int_0^a \rho' \, d\rho' \int_0^{2\pi} d\phi' \, e^{-ik\rho'\sin\theta\cos(\phi-\phi')} \tag{40}$$

First we shift variables $\phi - \phi' \to \phi$. Since we're just integrating around a circle, we can still take the limits of integration to be 0 to 2π. We'll also lump the additional variables into $x = k\rho' \sin\theta$. You may not recognize the angular integral, but since this is essentially a cylindrical problem, you might guess that it has something to do with Bessel functions. At this point it would be a good idea to open up your favorite book on special functions. It turns out the ϕ integral is a Bessel function

$$J_0(x) = \frac{1}{2\pi}\int_0^{2\pi} d\phi'\, e^{-ix\cos\phi}. \tag{41}$$

Thus, the electric field is

$$\begin{aligned}\mathbf{E}(\mathbf{x}) &= ikE_0(\hat{\mathbf{n}} \times \hat{\mathbf{y}})\frac{e^{ikr}}{r}\int_0^a \rho' J_0(k\rho'\sin\theta)\, d\rho' &(42)\\ &= ikE_0[\hat{\mathbf{n}} \times \hat{\mathbf{y}}]\frac{e^{ikr}}{r}\frac{1}{k^2\sin^2\theta}\int_0^{ka\sin\theta} xJ_0(x)\, dx &(43)\end{aligned}$$

Now we refer to our book on Bessel functions and find the recursive relation

$$xJ_0(x) = xJ_1'(x) + J_1(x) = \frac{d}{dx}(xJ_1). \tag{44}$$

Plugging this in and integrating yields

$$\mathbf{E}(\mathbf{x}) = ika^2E_0(\hat{\mathbf{n}} \times \hat{\mathbf{y}})\frac{e^{ikr}}{r}\frac{J_1(ka\sin\theta)}{ka\sin\theta}. \tag{45}$$

Now we'll find the power. We need to know

$$|\hat{\mathbf{n}} \times \hat{\mathbf{y}}|^2 = (\hat{\mathbf{n}} \cdot \hat{\mathbf{n}})(\hat{\mathbf{y}} \cdot \hat{\mathbf{y}}) - (\hat{\mathbf{n}} \cdot \hat{\mathbf{y}})^2 = 1 - (\hat{\mathbf{n}} \cdot \hat{\mathbf{y}})^2 = 1 - \sin^2\theta\sin^2\phi \tag{46}$$

There's ϕ dependence! This is due to the polarization of the electric field. For most of the intensity, however, $\theta \sim 0$, so that term isn't too big. We normalize this by the power incident on the aperture $P_i = (cE_0^2/8\pi) \times \pi a^2$. Then

$$\frac{1}{P_i}\frac{dP}{d\Omega} = \frac{(ka)^2}{\pi}\left(1 - \sin^2\theta\sin^2\phi\right)\left[\frac{J_1(ka\sin\theta)}{ka\sin\theta}\right]^2. \tag{47}$$

We plot the diffraction pattern below for $\phi = 0$ and $\phi = \pi/2$. Clearly the ϕ dependence makes very little difference. We'd like to know what the zeros are—measuring these could tell us the frequency of our light or the size of the aperture, whichever is unknown. Looking at our book of special functions, the first zero of J_1 occurs at $ka\sin\theta \simeq ka\theta = 3.832$. What about the envelope of the pattern? At large x,

$$\frac{J_1(x)}{x} \to \sqrt{\frac{2}{\pi x^3}}\cos\left(x - \frac{3\pi}{4}\right), \tag{48}$$

so the envelope falls out as $1/\theta^3$. At large angle x, the zeros are determined by the equation $\cos(ka\theta - 3\pi/4) = 0$, i.e. the zeros are evenly spaced.

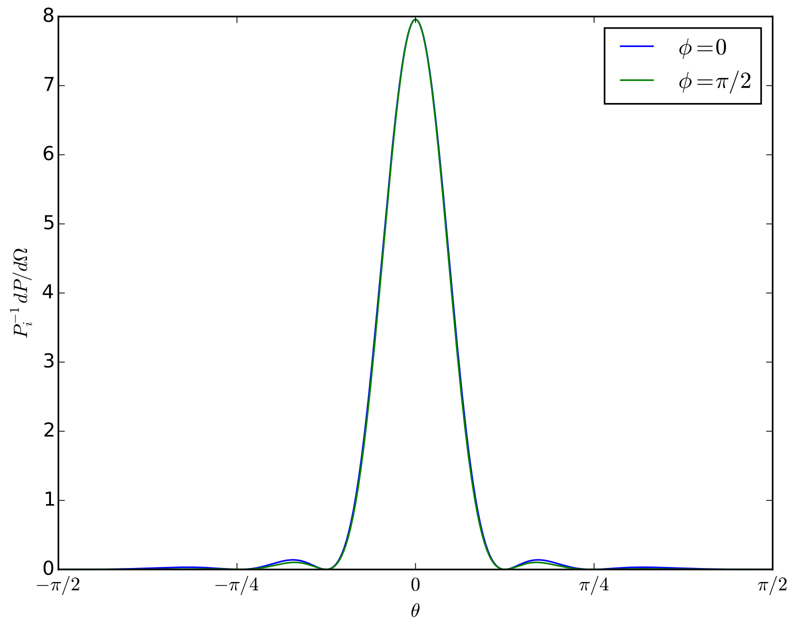


Figure 3: Diffraction from circular aperture for $ka = 10$.

We can find the transmission coefficient as well.

$$T = \int_0^{2\pi} \int_0^{\pi/2} \frac{1}{P_i} \frac{dP}{d\Omega} \sin\theta \, d\theta \, d\phi \tag{49}$$

$$= \frac{(ka)^2}{\pi} \int_0^{2\pi} \int_0^{\pi/2} \sin\theta \big(1 - \sin^2\theta \sin^2\phi\big) \left[\frac{J_1(ka\sin\theta)}{ka\sin\theta}\right]^2 d\theta \, d\phi \,. \tag{50}$$

This is daunting, but at least the ϕ integral is easy:

$$T = 2(ka)^2 \int_0^{\pi/2} \sin\theta \left(1 - \frac{1}{2}\sin^2\theta\right) \left[\frac{J_1(ka\sin\theta)}{ka\sin\theta}\right]^2 d\theta \,. \tag{51}$$

There's a nice limit when we take $ka \gg 1$, so the integrand is only sizable for $\theta \ll 1$. Then

$$T = 2(ka)^2 \int_0^{\pi/2} \theta \frac{J_1^2(ka\theta)}{(ka\theta)^2} \, d\theta = 2 \int_0^{ka\theta_{\max}} \frac{J_1^2(x)}{x} \, dx \,, \tag{52}$$

for some choice of $\theta_{\max}$. Using our Bessel function identity $xJ_0 = J_1 + xJ_1'$ from earlier, we can write

$$\frac{J_1(x)^2}{x} = J_1(x)[J_0(x) - J_1'(x)] = \frac{1}{2}\frac{d}{dx}\big[-J_0^2(x) - J_1^2(x)\big], \tag{53}$$

where the second step follows since $J_1 = -J_0'$. Then

$$T = 1 - J_0^2(ka\theta_{\max}) - J_1^2(ka\theta_{\max}). \tag{54}$$

What should we choose for $\theta_{\max}$? Well, it should certainly include the whole first peak. But after that, it doesn't really matter since the envelope goes like $1/\theta^3$. In the $ka \to \infty$ limit, $T \to 1$ since the Bessel functions go to zero. In the $ka \ll 1$ limit, Jackson shows that $T \propto (ka)^2$. Thus, as we shrink the whole, we get less transmission. But we have to be careful in the small hole limit because the field at the aperture will become perturbed, so we can no longer use $\hat{\boldsymbol{\epsilon}}_0 E_0 e^{i\mathbf{k}_0\cdot\mathbf{x}'}$ as the field in the Smythe-Kirchhoff integral.

4.4 Additional Topics in a Nutshell

Here we briefly introduce ***Babinet's principle*** and the ***optical theorm*** without going into any serious detail.

Babinet's principle

Let's restrict our discussion to scalar diffraction for simplicity. The Kirchhoff formula is

$$\psi(\mathbf{x}) = \frac{1}{4\pi}\int_{S_1} \hat{\mathbf{n}}'\cdot\left[\boldsymbol{\nabla}'\psi + ik\psi\frac{\mathbf{R}}{R}\right]\frac{e^{ikR}}{R}\,da'\,. \tag{55}$$

where S_1 is the entire flat surface, including screen and aperture. We can break S_1 into two regions: S_a, which contains everything but the hole, and S_b, which contains hole. Clearly, the field $\psi(\mathbf{x})$ can be expanded as

$$\psi = \psi_a + \psi_b, \tag{56}$$

where ψ_a and ψ_b are obtained by restricting the integral in the Kirchhoff formula to S_a and S_b. Consider sending in an incident wave $\psi(\mathbf{x}')$. If the entire surface S_1 is aperture, then ψ_b is the diffracted wave we would get if S_b was empty and S_a was solid (i.e. the hole was indeed a hole and the screen was indeed a screen), while ψ_a is the diffracted wave we would get if S_b was solid and S_a was empty (i.e. the hole was filled in and the screen was taken away). But if S_1 is entirely aperture, then there is no diffracted wave $\psi^{\text{diff}}(\mathbf{x}) = 0$. Thus,

$$\psi_a^{\text{diff}} = -\psi_b^{\text{diff}}. \tag{57}$$

In particular, the intensities are equal

$$|\psi_a|^2 = |\psi_b|^2 = I_{\text{diff,a}} = I_{\text{diff,b}}. \tag{58}$$

Thus, the diffraction patterns from complimentary screens are equal. This is ***Babinet's principle***.

As usual, when we consider vectors, there are additional complications as the polarization creates additional effects. Essentially, the scattering for a magnetic field of a given polarization off an aperture is equal to the scattering of an electric field of the same polarization off the complementary screen.

Optical Theorem

We defined the scattering amplitude $\mathbf{f}(\mathbf{k}_0, \mathbf{k})$ by

$$\hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{E}_{\text{sc}} = \frac{e^{ikr}}{r} \hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{f}(\mathbf{k}_0, \mathbf{k}) E_0, \tag{59}$$

which yields the differential cross section via

$$\frac{d\sigma}{d\Omega} = |\hat{\boldsymbol{\epsilon}}^* \cdot \mathbf{f}|^2. \tag{60}$$

The optical theorem says that there's a relation between the total cross section and the scattering amplitude for forward scattering:

$$\sigma_{\text{tot}} = \frac{4\pi}{k} \operatorname{Im}\left[\hat{\boldsymbol{\epsilon}}_0^* \cdot \mathbf{f}(\mathbf{k}_0, \mathbf{k} = \mathbf{k}_0)\right]. \tag{61}$$

Here, the total cross section σ_{tot} consists of the total elastic/scattering cross section and the total absorption cross section

$$\sigma_{\text{tot}} = \sigma_{\text{elastic}} + \sigma_{\text{abs}}. \tag{62}$$

The way one proves this is by brute force. One can use the vector Smythe-Kirchhoff formula where the surface S_1 is a small surface around the scattering center and S_2 is a large sphere out at infinity. This gives σ_{elastic}. The absorption cross section is

$$\sigma_{\text{abs}} = -\frac{c}{8\pi} \int_{S_1} \hat{\mathbf{n}}' \cdot \operatorname{Re}\left(\mathbf{E} \times \mathbf{B}^*\right) da' \,, \tag{63}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the total fields, incident plus scattered.

We won't do the proofs. In fact, they're nicer in quantum mechanics, where one has unitarity. But the result itself is nice because it's exact. Sometimes you can choose parametrizations that encode such an exact result, and this makes one's life easier.

5 Special Relativity

Einstein said it best. In his original 1905 paper on special relativity, he wrote

> ...the same laws of electrodynamics and optics will be valid for all frames of reference for which the equations of mechanics hold good. We will raise this conjecture (the purport of which will hereafter be called the Principle of Relativity) to the status of a postulate, and also introduce another postulate, which is only apparently irreconcilable with the former, namely, that light is always propagated in empty space with a definite velocity c which is independent of the state of motion of the emitting body.

The rest is technical details.

People noticed that Maxwell's equations were invariant under Lorentz transformations. This comes down to the fact that the d'Alembertian operator is Lorentz invariant. People also deduced how the electric and magnetic fields transform under change of reference frame (although they do not transform like 4-vectors, since together the electric and magnetic fields form a tensor). However, it was Einstein was the one who realized that special relativity was not specific to electricity and magnetism, but was a property of spacetime itself. Nowadays, the procedure is run backwards. We begin by postulating Lorentz symmetry and internal (e.g. gauge) symmetries, and derive the equations of motion from there.

Let's go back to Einstein's quote. The two principles are

1. The Laws of Nature are the same in all inertial frames,
2. The speed of light is independent of the motion of the source.

Two frames are inertial with respect to each other if, for example, the origin of the second frame is at $\mathbf{r}_0 + \mathbf{v}(t - t_0)$ as seen by the first frame, where $\mathbf{v}$ is constant. That the speed of light is independent of the source means that $c = c$ in all inertial frames.

Still more generally, consider two events separated by Δx in space and Δt in time, as seen by an observer in one inertial frame. These events are separated by $\Delta x'$ and $\Delta t'$ by another observer. We can begin relativity with the statement

$$(\Delta s)^2 = c^2(\Delta t)^2 - (\Delta x)^2 = c^2(\Delta t')^2 - (\Delta x')^2, \tag{1}$$

i.e. the quantity $(\Delta s)^2$ is the same in all inertial frames. In the special case of light,

$$c^2(\Delta t)^2 - (\Delta x)^2 = 0. \tag{2}$$

This is the fundamental symmetry of space and time from which special relativity arises.

This should be contrasted with Galilean invariance, which relates the coordinates of two frames by $x' = x + vt$ and $t' = t$. The quantity

$$R^2 = (\mathbf{x}_1 - \mathbf{x}_2)^2 = (\Delta x)^2 = (\Delta x')^2 \tag{3}$$

is invariant, and so is Δt. Note that Newton's second law exhibits Galilean invariance

$$\mathbf{F} = m\frac{d^2\mathbf{x}}{dt^2} = m\frac{d^2\mathbf{x}'}{dt'^2}. \tag{4}$$

However, we know that Lorentz symmetry is the correct symmetry of spacetime, so Newton's second law can only be approximate.

5.1 Lorentz Transformations

In our introduction we stated that we could begin special relativity with the invariance of the spacetime interval

$$(\Delta s)^2 = c^2(\Delta t)^2 - (\Delta x)^2. \tag{5}$$

The fact that $(\Delta s)^2 = 0$ for two points on a photon's trajectory implies that c is constant. Actually, we could have obtained the constancy of c even if the spacetime interval transformed as

$$(\Delta s_1)^2 = \lambda(v^2)(\Delta s_2)^2 \tag{6}$$

for some function $\lambda(v^2)$. With certain homogeneity and isotropy assumptions on spacetime, one can show that a function λ that depends only on v^2 is indeed the only option. Then the requirement that two successive Lorentz transformations yield another Lorentz transformation implies that $\lambda(v^2) \equiv 1$. We won't go into the details here, but suffice it to say that we can derive the invariance of the spacetime interval (5) from weaker assumptions.

We want to consider the coordinates x, t in one frame and relate them to the coordinates x', t' in another frame using the invariant spacetime interval. Minkowski's analogy is to consider the Euclidean invariant

$$R^2 = x^2 + y^2 + z^2 = x'^2 + y'^2 + z'^2. \tag{7}$$

If we specialize to fixing one dimension $z = z'$, then we know the transformations that leave R^2 invariant; they're just rotations. To find how to relate coordinates with the invariant $x^2 - c^2t^2$, we simply make the substitution $w = it$

$$x^2 - c^2t^2 = x^2 + c^2w^2, \tag{8}$$

This is a Euclidean interval, which we preserve with rotations:

$$cw' = cw\cos\theta + x\sin\theta, \tag{9}$$

$$x' = x\cos\theta - cw\sin\theta. \tag{10}$$

Substituting back in $w = it$ and using $i\sin\theta = \sinh i\theta$ and $\cos i\theta = \cosh\theta$ (write these functions in terms of exponentials to see this), we obtain

$$ct' = (-i)\left[ict\cos\left(i^2\theta\right) + x\sin\theta\right] = ct\cosh\eta - x\sinh\eta, \tag{11}$$

$$x' = x\cos\left(i^2\theta\right) - ict\sin\theta = x\cosh\eta - ct\sinh\eta. \tag{12}$$

Where $\eta = i\theta$ is called the ***rapidity***, and is a *real* variable. In matrix form, the transformation is

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \cosh\eta & -\sinh\eta \\ -\sinh\eta & \cosh\eta \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix}. \tag{13}$$

Physically, what is η? Consider two frames S and S' in which the origin of S' at $x' = 0$ moves with velocity v along the the x-axis of frame S. The location of the origin of S' as seen from S is $x = vt$. So the Lorentz transformation gives

$$x' = 0 = vt\cosh\eta - ct\sin\eta, \tag{14}$$

which implies

$$\tanh \eta = \frac{v}{c} \equiv \beta. \tag{15}$$

Using the identity $\cosh^2 \eta - \sin^2 \eta = 1$, one can derive

$$\cosh \eta = \frac{1}{\sqrt{1-\beta^2}} \equiv \gamma, \tag{16}$$

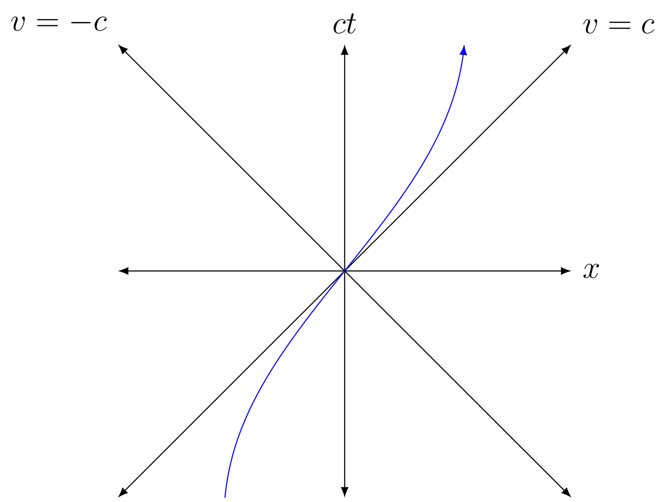
$$\sinh \eta = \frac{\beta}{\sqrt{1-\beta^2}} = \beta\gamma \tag{17}$$

Thus, in more familiar notation, the Lorentz tranformation is

$$\begin{pmatrix} ct' \\ x' \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \begin{pmatrix} ct \\ x \end{pmatrix} \tag{18}$$

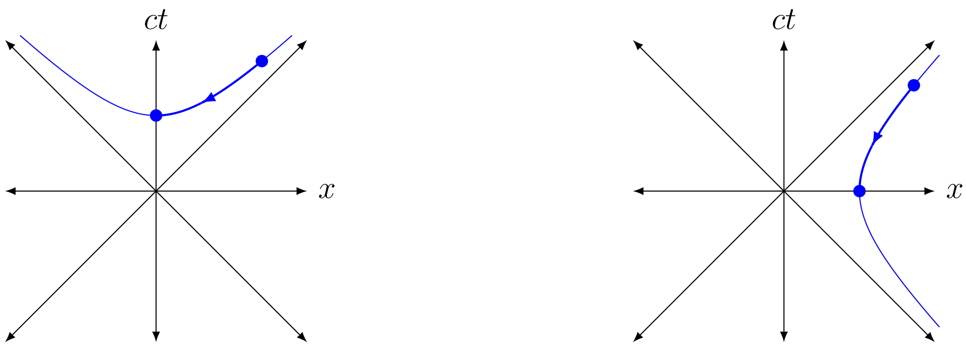
for a boost along x. The other spatial coordinates remain unchanged; $y' = y$ and $z' = z$.

Spacetime diagrams like the one below are often useful in special relativity.



Light travels along the diagonals. Imagine rotating the diagram into a second space dimension (and then into a third!), then diagonal lines becomes a cone, called the ***light cone***. Physical objects travel at speeds less than the speed of light, and hence travel within the light cone. Such a trajectory is shown in blue on the diagram.

This diagram is drawn in a certain frame. What if we transform to another frame? We know $x^2 - c^2t^2$ is invariant, and lines with $s^2 = x^2 - c^2t^2$ for some constant s^2 are hyperbolas. Thus, points on a spacetime diagram move along hyperbolas under Lorentz transformation, as displayed in the two diagrams below.



(a) (b)

Consider the spacetime interval between two events

$$s_{12}^2 = c^2(t_1 - t_2)^2 - (\mathbf{x}_1 - \mathbf{x}_2)^2. \tag{19}$$

If $s_{12}^2 > 0$, the events are said to be ***timelike separated***, and we define the proper time $c\Delta\tau \equiv s_{12}$. Note that the proper time is an invariant. Plot (a) above displays an event timelike separated from the origin. We can boost into a frame where these events happen at the same location in space, shown by the arrow along the hyperbola in plot (a); in this frame the proper time is just $\Delta\tau = \Delta t$. The hyperbola makes clear that the proper time is the shortest perceivable time between the events. The hyperbola also shows that two events that are timelike separated have a clear notion of which came first, hence of cause and effect.

Consider instead two events such that $s_{12}^2 < 0$, called ***spacelike separated*** events. A point spacelike separated with respect to the origin is displayed in plot (b). Moving the point along the hyperbola, we see that there's a frame in which they occur at the same point in time, and there are also reference frames where either event occurs before the other. The time ordering is not unique, so it's not possible to exchange information between spacelike separated points—to do so would violate causality.

Here is another argument that $\Delta\tau$ is the shortest time between timelike separated points and a consequence thereof. Imagine a particle has a lifetime τ in its rest frame, hence τ is its proper time. Consider a frame in which the particle has velocity $u(t)$, so $d\mathbf{x} = \mathbf{u}dt$. Then

$$c^2 d\tau^2 = ds^2 = c^2(dt)^2 - |d\mathbf{x}|^2 = (dt)^2\big(c^2 - u^2\big) = \frac{c^2}{\gamma^2}(dt)^2. \tag{20}$$

We take a square root and write

$$cd\tau = ds = \frac{c\,dt}{\gamma(t)}. \tag{21}$$

Then

$$\Delta t = \int_{s_1}^{s_2} \gamma(t)\frac{ds}{c} \geq \int_{\tau_1}^{\tau_2} d\tau = \Delta\tau \tag{22}$$

where the inequality follows since $\gamma > 1$. Taking the interval of integration to be the whole lifetime of the particle, this tells us that the particle lives longer in a frame where it's moving. This is time dilation in action.

5.2 4-Vectors

A ***4-vector*** is any quantity which transforms under a Lorentz transformation like $x = (ct, \mathbf{x})$. We write it like

$$A \equiv (A_0, A_1, A_2, A_3), \tag{23}$$

where A_0 is like ct, and $\mathbf{A} = (A_1, A_2, A_3)$ transform like the ordinary 3-vector $\mathbf{x}$. There's a convention that four-vectors have their indices as superscripts instead of subscripts; we're ignoring that convention for now. The zero component transforms like

$$A_0' = \gamma(A_0 - \boldsymbol{\beta} \cdot \mathbf{A}) \tag{24}$$

under Lorentz transformation. The components parallel and perpendicular to the boost direction transform like

$$A_\parallel' = \gamma\big(A_\parallel - \beta A_0\big) \tag{25}$$

$$A_\perp' = A_\perp \tag{26}$$

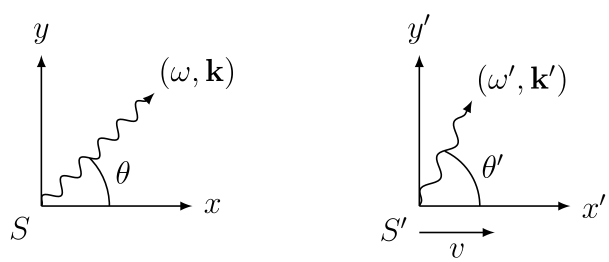
The scalar product of two four vectors is invariant

$$A \cdot B = A_0 B_0 - \mathbf{A} \cdot \mathbf{B}. \tag{27}$$

To prove this, one can just substitute in the Lorentz transformations of A and B and turn the crank.

Relativistic Doppler Shift

We want to relate the frequency/wavelength of light in one frame with its frequency/wavelength in another frame. We'll discover that ω and $\mathbf{k}$ transform like a four vector. Consider boosting in a direction at an angle θ with the boost vector, as shown below.



Take the light to be a pulse with ϕ wave crests. The number of crests ought to be the same in both frames, so ϕ is an invariant. But

$$e^{i\phi} = e^{i(\omega t - \mathbf{k}\cdot\mathbf{x})} = e^{i(\omega' t' - \mathbf{k}'\cdot\mathbf{x}')}. \tag{28}$$

We know $(ct, \mathbf{x})$ transforms like a four-vector and since ϕ is an invariant, it must be that $k = (\omega/c, \mathbf{k})$ transforms like a four-vector too. In general,

$$\omega' = \gamma(\omega - \boldsymbol{\beta} \cdot \mathbf{k}) = \gamma\omega(1 - \beta\cos\theta), \tag{29}$$

where we have used $\omega = c|\mathbf{k}|$. If the boost is in the direction of the wavevector $\mathbf{k}$, so that the source is moving away, then

$$\omega' = \omega\gamma(1 - \beta) = \omega\sqrt{\frac{1-\beta}{1+\beta}} < \omega, \tag{30}$$

i.e., we get a redshift. Boosting in the opposite direction gives

$$\omega' = \omega\gamma(1 + \beta) = \omega\sqrt{\frac{1+\beta}{1-\beta}} > \omega, \tag{31}$$

so we get a blueshift when the source is moving towards us. Note that we'd also get a Doppler shift if we boosted perpendicular to the wave vector. In this case $\omega' = \gamma\omega$.

The wavevector direction also changes:

$$\frac{k'_\perp}{k'_\parallel} = \tan\theta' = \frac{1}{\gamma}\frac{|\mathbf{k}|\sin\theta}{|\mathbf{k}|\cos\theta - \beta|\mathbf{k}|} = \frac{\sin\theta}{\gamma(\cos\theta - \beta)}. \tag{32}$$

The passes through $\theta' = \pi/2$ when $\beta = \cos\theta$ and goes to π as $\gamma \to \infty$.

Gravitational Redshift

There is another kind of redshift—the gravitational redshift—that is relevant for the kind of precision measurements that get done at JILA. Imagine we have a tower of height h. If we shoot light from the top to bottom, it suffers a shift in frequency

$$\frac{\Delta\omega}{\omega} = -\frac{gh}{c^2}, \tag{33}$$

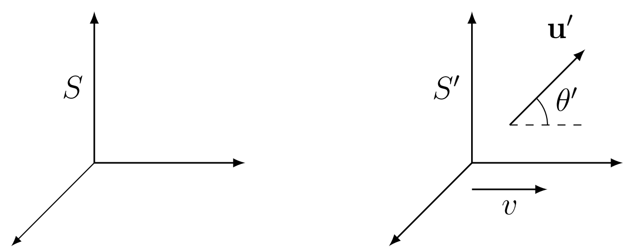
which we could guess just by dimensional analysis. For a precision measurement, we can imagine $h = 0.1\,\mathrm{m}$. Then we get

$$\frac{\Delta\omega}{\omega} = 10^{-17}. \tag{34}$$

Detection of such a redshift is not unobtainable, especially at JILA! Note that this is not a Doppler shift; the emitter and receiver are at rest. The redshift occurs because we are not in an inertial frame! Take a general relativity course if this sounds interesting.

More 4-vectors

So far we have constructed $(ct, \mathbf{x})$ and $(\omega/c, \mathbf{k})$ as 4-vectors. Let us construct some more. Consider two frames S and S', where S' is moving at velocity v relative to S. Consider an object moving at velocity $\mathbf{u}'$ which makes an angle θ' with $\mathbf{x}'$ as seen in S'. What is the object's velocity $\mathbf{u}$ as seen in S?



We do a Lorentz transformation on dx to obtain

$$
\begin{aligned}
dx_0 &= \gamma_v(dx_0' + \beta dx_1') && (35)\\
dx_1 &= \gamma_v(dx_1' + \beta dx_0') && (36)\\
dx_2 &= dx_2' && (37)\\
dx_3 &= dx_3', && (38)
\end{aligned}
$$

where $x_0 = ct$ and $\gamma_v = 1/\sqrt{1 - v^2/c^2}$. We can take ratios of the differentials to get the velocity in S:

$$u_\parallel = c\frac{dx_1}{dx_0} = c\frac{dx_1' + \beta dx_0'}{dx_0' + \beta dx_1'} = \frac{u_\parallel' + v}{1 + vu_\parallel'/c^2}. \tag{39}$$

$$u_\perp = c\frac{dx_2}{dx_0} = c\frac{dx_2'}{\gamma_v(dx_0' + \beta dx_1')} = \frac{u_\perp'}{\gamma_v(1 + vu_\parallel'/c^2)}. \tag{40}$$

The first result is the well-known addition of velocities law for the velocity along the boost direction. The second result is the less-well-known addition of velocities law for a velocity perpendicular to the boost direction. Another way to do this would be to perform successive Lorentz transformations and use hyperbolic trig identities to write the result in the typical form of a Lorentz transformation.

What is interesting is that this shows that the velocity is ***not*** part of a 4-vector. But there is a related 4-vector, called the 4-velocity, which is a 4-vector. This is the derivative U of the 4-vector x with respect to the proper time. The position x is a 4-vector, the proper time is an invariant, so this derivative will automatically be a 4-vector. Recall that ordinary coordinate time is $t = \gamma\tau$. Then

$$U_0 = \frac{dx_0}{d\tau} = \frac{dt}{d\tau}\frac{dx_0}{dt} = \gamma_u c \tag{41}$$

$$\mathbf{U} = \frac{d\mathbf{x}}{d\tau} = \frac{dt}{d\tau}\frac{d\mathbf{x}}{dt} = \gamma_u \mathbf{u} \tag{42}$$

Let's check that the length is invariant, as it should be for any four-vector. We compute

$$U_0^2 - \mathbf{U}^2 = \gamma_u^2\big[c^2 - u^2\big] = c^2, \tag{43}$$

which is of course invariant by the second postulate of relativity! Note that $\mathbf{U}$ reduces to the velocity when $\gamma_u \to 1$.

The 4-vector transformation of U is consistent with the addition of velocities formulae we derived above. To show this, we need the identity

$$\gamma_u = \gamma_v \gamma_{u'} \left(1 + \frac{\mathbf{v}\cdot\mathbf{u}'}{c^2}\right), \tag{44}$$

where $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{u}'$ are defined as before. This can be proven by decomposing u into parallel and perpendicular parts and substituting the addition of velocities formulae. Then

$$\gamma_u u_\parallel = \gamma_v \gamma_{u'} \left(1 + \frac{v u'_\parallel}{c^2}\right) \frac{u'_\parallel + v}{1 + v u'_\parallel / c^2} = \gamma_v\big(\gamma_{u'} u'_\parallel + \beta \gamma_{u'} c\big) \tag{45}$$

$$\gamma_u u_\perp = \gamma_{u'} \gamma_v \left(1 + \frac{v u'_\parallel}{c^2}\right) \frac{u'_\perp}{\gamma_v (1 + v u'_\parallel / c^2)} = \gamma_{u'} u'_\perp, \tag{46}$$

which is the Lorentz transformation of the spatial components of U.

Thus, we add $U = (c\gamma_u, \mathbf{u}\gamma_u)$ to our list of 4-vectors. For an observer at rest, we have $U = (c, 0, 0, 0)$, which is quite useful! If we have an arbitrary four-vector $A = (A_0, A_1, A_2, A_3)$, we can get its timelike component as measured in our rest frame by dotting it with U:

$$A \cdot U = cA_0. \tag{47}$$

This is sometimes a useful trick to know.

We can construct the 4-acceleration in the same way we constructed the 4-velocity. We just differentiate again with respect to proper time.

$$A = \frac{dU}{d\tau} = \left[\gamma \frac{d}{dt}(\gamma c), \gamma \frac{d}{dt}(\gamma \mathbf{u})\right] \tag{48}$$

We compute

$$\frac{d\gamma}{dt} = \frac{d}{dt} \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} = \frac{\mathbf{u}\cdot\mathbf{a}}{c^2}\gamma^3, \tag{49}$$

where $\mathbf{a} = d\mathbf{u}/dt$. Plugging this into A yields

$$A = \left(\gamma^4 \frac{\mathbf{u}\cdot\mathbf{a}}{c}, \gamma^4 \frac{\mathbf{u}\cdot\mathbf{a}}{c^2}\mathbf{u} + \gamma^2 \mathbf{a}\right). \tag{50}$$

In our rest, frame, this reduces to

$$A = (0, \mathbf{a}). \tag{51}$$

Since $U = (c, \mathbf{0})$ in this frame its clear that $A \cdot U = 0$, which is then true in all frames.

Why should we care to write A down like this? We want to come up with relativistic dynamics as a revision to Newtonian mechanics. Given that **a** is important in Newtonian mechanics and our relativistic revision must have this as a limit, its a good guess that A will be a good relativistic generalization. We'll show that it is eventually.

Let's go after momentum now. It's hard to find a satisfying "derivation" of relativistic momentum. It's easy to write a relativistic generalization of energy and momentum that reduces to Newtonian energy and momentum. But it's a question that must be decided by experiment (or more high-brow mathematics) whether this generalization is useful or physical.

We'll motivate the 4-vector momentum with some insight from quantum mechanics. Consider a wavevector $\psi(t, \mathbf{x})$. If we infinitesimally translate in space we get

$$\psi(\mathbf{x} + \delta\mathbf{x}) = \psi(\mathbf{x}) + \delta\mathbf{x} \cdot \nabla\psi = \hat{S}\psi(q), \tag{52}$$

where

$$\hat{S} = 1 + \delta\mathbf{q} \cdot \nabla = 1 + \frac{i}{\hbar}\delta\mathbf{q} \cdot \mathbf{p} \tag{53}$$

This tells us that momentum is the generator of spatial translations. We play the same game for time

$$\psi(t + \delta t) = \psi(t) + \delta t \frac{\partial\psi}{\partial t} = \hat{U}\psi(t) \tag{54}$$

where

$$U = 1 - \frac{i}{\hbar}\delta t H; \tag{55}$$

we say the Hamiltonian is the generator of time translations. In a nonrelativistic quantum mechanics class, H and P have nothing to do with each other. But relativity says that dt and dx are related by Lorentz transformations. Thus, energy and momentum must be related as well.

Let's consider a free particle. We know for nonrelativistic free particles

$$\mathbf{p} = m\mathbf{u}, \qquad E = E(0) + \frac{p^2}{2m} \tag{56}$$

These are interesting quantities because they're conserved. We better get these formulae as nonrelativistic limits for our relativistic generalizations. According to Einstein, the only possible generalization of energy and momentum into a four vector involves the four velocity, i.e.,

$$\mathbf{p} = m\frac{d\mathbf{x}}{dt} \tag{57}$$

"must" generalize to

$$P = m\frac{dx}{d\tau} = mU. \tag{58}$$

The reason we say "must" is because it is the only four-vector whose nonrelativistic limit is $(E(0) + m\mathbf{u}^2/2, m\mathbf{u})$. Of course this applies fo $m \neq 0$. Redefining $E = \gamma mc^2$ and $\mathbf{p} = \gamma m\mathbf{u}$, we have

$$P = \left(\frac{E}{c}, \mathbf{p}\right). \tag{59}$$

In the limit $u^2 \ll c^2$ we recover

$$E = \frac{mc^2}{\sqrt{1 - u^2/c^2}} = mc^2 + \frac{1}{2}mu^2 + \cdots \tag{60}$$

and $\mathbf{p} = m\mathbf{u}$, as required.

Alternatively, here's a "joke" derivation of $P = (\gamma mc, \gamma m\mathbf{u})$. We invoke the de Broglie relations $\mathbf{p} = \hbar\mathbf{k}$ and $E = \hbar\omega$. Since we know $(\omega/c, \mathbf{k})$ is a four-vector, we know

$$P = \left(\frac{\hbar\omega}{c}, \hbar\mathbf{k}\right) = \left(\frac{E}{c}, \mathbf{p}\right)$$

is a four-vector.

We'll assume that our new relativistic E and $\mathbf{p}$ are conserved. Once we assume this in one frame, we immediately get it in all frames. To see this, consider some elastic process $a + b \to c + d + e + \cdots$. We assume that in a particular frame

$$\sum \mathbf{p}_i \equiv \sum_{\text{initial}} \mathbf{p}_j - \sum_{\text{final}} \mathbf{p}_k = 0, \tag{61}$$

where we've redefined p_i just by flipping the signs on the final p_i. If we also assume energy conservation in this frame

$$\sum E_i \equiv \sum E_j - \sum E_k = 0. \tag{62}$$

Now when we transform to a new frame, $(E/c, \mathbf{p})$ transforms like a 4-vector, so we get

$$\sum E_i' = \gamma\Big(\sum E_i - \beta c \sum p_{\|,i}\Big) = 0 \tag{63}$$

$$\sum p_{\|,i}' = \gamma\Big(\sum p_{\|,i} - \frac{\beta}{c}\sum E_i\Big) = 0 \tag{64}$$

$$\sum p_{\perp i}' = \sum p_{\perp,i} = 0 \tag{65}$$

Thus, energy-momentum is conserved in all frames.

The fact that P is a four-vector implies that $E^2/c^2 - p^2$ is invariant. Using $P = (\gamma mc, \gamma m\mathbf{v})$, this is

$$E^2 - p^2c^2 = m^2c^2\gamma^2\big(c^2 - v^2\big) = (mc^2)^2, \tag{66}$$

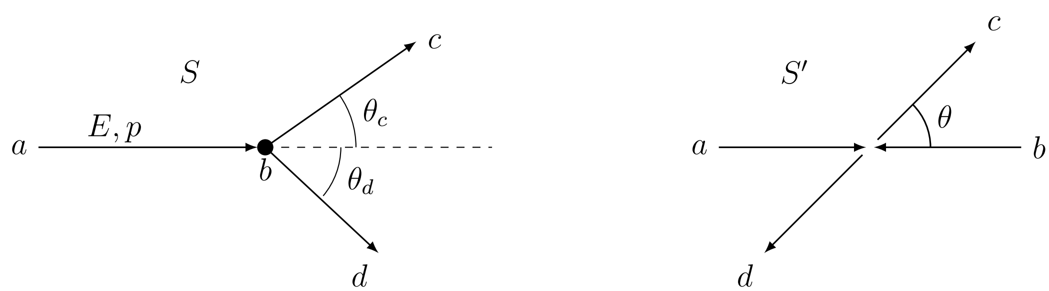
which is the familiar relativistic energy-momentum relation. Of course, the rest mass energy of a particle is $E = mc^2$. A few more useful formulae come from comparing the forms $(E/c, \mathbf{p})$ and $(\gamma mc, \gamma m\mathbf{u})$:

$$\frac{E}{mc^2} = \gamma, \qquad \frac{cp}{E} = \frac{\gamma mv}{\gamma mc} = \beta. \tag{67}$$

If we know two out of m, E, and p, we can find the velocity v using the above relations.

One set of practical problems in relativity includes kinematics in different frames, e.g. a center of mass frame and a lab frame. One way to get between different frames is, of course, via Lorentz transformation. But a nicer way is often to use invariants, like products of 4-vectors. Invariants are nice because, well, they're invariant.

Example 1. Consider the kinematics of an elastic collision $a + b \to c + d$ in a frame with a fixed target and in the CM frame. We'll set $c = 1$ here and in the future. Let the particles have equal mass m.



(a) Fixed target frame. (b) Center of mass frame.

From the diagrams, we can see

$$P_a = (E, p, 0, 0) \tag{68}$$

$$P_b = (m, 0, 0, 0) \tag{69}$$

$$P_c = (E_c, p_c \cos\theta_c, p_c \sin\theta_c, 0) \tag{70}$$

$$p_d = (E_d, p_d \cos\theta_d, -p_d \sin\theta_d, 0) \tag{71}$$

In the center of mass frame, things are simpler. Particles a and b come at each other with equal and opposite momentum. Since they have the same mass they must also have the same energy by $E^2 = p^2 + m^2$. Thus,

$$P_{a'} = (E', p', 0, 0) \tag{72}$$

$$P_{b'} = (E', -p', 0, 0) \tag{73}$$

$$P_{c'} = (E', p' \cos\theta, p' \sin\theta, 0) \tag{74}$$

$$P_{d'} = (E', -p' \cos\theta, -p' \sin\theta, 0), \tag{75}$$

with $P_{c'}$ and $P_{d'}$ following by conservation of energy and momentum. What are the lab E, p, E_c, p_c in terms of θ, E'? We could figure out a boost from one frame to the next. But for some quantities, invariants are faster. For example, it's easy to set up the equation $(P_i \pm P_j)^2 = (P_{i'} \pm P_{j'})^2$. Setting $i = a$ and $j = b$,

$$(P_a + P_b)^2 = (P_{a'} + P_{b'})^2 \tag{76}$$

$$2m^2 + 2P_a \cdot P_b = 4E'^2 \tag{77}$$

$$2m^2 + 2mE = 4E'^2 \tag{78}$$

This relates the energies in the lab and center of mass frames. For the angles,

$$(P_a - P_c)^2 = (P_{a'} - P_{c'})^2 \tag{79}$$

$$2m^2 - 2[EE_c - pp_c \cos\theta_c] = 2m^2 - 2\big(E'^2 - p'^2 \cos\theta\big). \tag{80}$$

Likewise,

$$(P_b - P_c)^2 = (P_{b'} - P_{c'})^2 \tag{81}$$

$$2m^2 - 2mE_c = 2m^2 - 2\big(E'^2 + p'^2 \cos\theta\big). \tag{82}$$

These can be solved for the unknowns.

Another example is the detection of the Higgs. The Higgs decays into two photons, and we detect the photon energy and momenta. How do we know the photons came from a Higgs decay? We can sum the photons 4-momenta and square the sum; if the photons came from a Higgs the result should be m_H^2 by conservation of 4-momentum. If that's what we find, then we can claim we saw a Higgs decay.

5.3 Differential Geometry of Special Relativity

Ultimately, we want to construct Lagrangians whose dynamics are consistent with special relativity. What we need are objects that transform simply under Lorentz transformation, and a formalism that allows us to deal with Lorentz transformations automatically, in such a way that we don't have to think. In this section, we'll discuss the bare minimum of differential geometry that we need for special relativity. This will give us the desired formalism.

We describe spacetime in terms of a 4-dimensional space whose dimensions are indexed by $\mu = 0, 1, 2, 3$. We write 4-vectors as

$$x^\mu = \big(x^0, x^1, x^2, x^3\big). \tag{83}$$

The superscript is an index, not an exponent. For example $x^0 = ct$ and x^i is the ith component of $\mathbf{x}$. We will make a distinction between lower and upper indices, so the index placement will be important.

Imagine a coordinate transformation

$$x^\mu \to x^{\mu'}. \tag{84}$$

We want to consider observables which transform simply under this transformation. In particular, we want them to transform linearly amongst themselves. Such observables are generally defined as ***tensors of rank k***, which may depend on spacetime points. Rank 0 tensors are scalars. These are unchanged when we change coordinates or perform Lorentz transformations, i.e. a scalar s transforms like $s' = s$. Examples include masses and dot products of 4-vectors.

Rank 1 tensors are vectors, for example the spacetime coordinate x^μ. Then $x^{\mu'}$ is a linear function of x^μ, e.g.

$$x' = \gamma x + \beta\gamma t \qquad x = \gamma x' - \beta\gamma t' \tag{85}$$
$$t' = \gamma t + \beta\gamma x \qquad t = \gamma t' - \beta\gamma x' \tag{86}$$

Since the transformation is linear, we can write this as

$$x' = \left(\frac{\partial x'}{\partial x}\right) x + \left(\frac{\partial x'}{\partial t}\right) t \tag{87}$$
$$t' = \left(\frac{\partial t'}{\partial t}\right) t + \left(\frac{\partial t'}{\partial x}\right) x \tag{88}$$

In general, we have a new set of coordinates $x^{\mu'}$ related to the old coordinates by

$$x^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^\nu} x^\nu, \tag{89}$$

where we sum repeated indices. Note that we sum over an upper index in the denominator and an upper index in the numerator. Conversely, for an arbitrary collection of 16 numbers, i.e. a 4×4 matrix $\Lambda^{\mu'}{}_\nu$, you could define the coordinate transformation

$$x^{\mu'} \equiv \Lambda^{\mu'}{}_\nu x^\nu. \tag{90}$$

Now we sum over a lower index and an upper index, both in the numerator. In general, a lower index in the numerator acts like an upper index in denominator. The rule is to sum repeated upper and lower indices. It's a handy grammatical rule since this is the only context in which repeated indices should appear, e.g. two repeated upper indices should not appear with both in the numerator. A vector is by definition an object that transforms in the same way as the coordinates. If we have $A^\mu = (A^0, A^1, A^2, A^3)$, then in the new frame

$$A^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^\nu} A^\nu = \Lambda^{\mu'}{}_\nu A^\nu. \tag{91}$$

We can also think of vectors in a coordinate invariant way. We could choose a basis $\hat{e}_{(\mu)}$ for the space of 4-vectors and write a vector as a linear combination of basis vectors, like we always do in regular three dimensional physics:

$$V = V^\mu \hat{e}_{(\mu)} = V^{\nu'} \hat{e}_{(\nu')} = \Lambda^{\nu'}{}_\mu V^\mu \hat{e}_{(\nu')}. \tag{92}$$

Note that since the components V^μ are arbitrary, this tells us that $\hat{e}_{(\mu)} = \Lambda^{\nu'}{}_\mu \hat{e}_{(\nu')}$.

Given a transformation $\Lambda^{\mu'}{}_\nu$ from a frame S into a frame S', we define $\Lambda^\mu{}_{\nu'}$ to be the inverse transformation, i.e.

$$\Lambda^\mu{}_{\nu'} \Lambda^{\nu'}{}_\rho = \delta^\mu{}_\rho \quad \text{and} \quad \Lambda^{\mu'}{}_\nu \Lambda^\nu{}_{\rho'} = \delta^{\mu'}{}_{\rho'}. \tag{93}$$

There's another kind of object that carries one index. These are called either covariant 4-vectors, dual vectors, or one-forms depending on what generation you're from and how much mathematics you know. I'll try to stick to calling them dual vectors. We write these as B_μ and define them by the transformation law

$$B_{\mu'} = \frac{\partial x^\nu}{\partial x^{\mu'}} B_\nu = \Lambda^\nu{}_{\mu'} B_\nu. \tag{94}$$

The transformation law is similar to the vector transformation law, but with the indices on the transformation switched. In coordinate invariant form we can choose a basis $\theta^{(\mu)}$ for the dual vector space and write

$$B = B_\mu \hat{\theta}^{(\mu)}. \tag{95}$$

In principle, a vector $V = V^\mu \hat{e}_{(\mu)}$ and the above dual vector have nothing to do with each other, but typically we choose basis vectors so that

$$\hat{\theta}^{(\nu)} \hat{e}_{(\mu)} = \delta^\nu{}_\mu. \tag{96}$$

The dual vector space is the space of linear functions from vectors to real numbers, so if you wish the right hand side above can be thought of as the function $\hat{\theta}^{(\nu)}$ acting on the vector $\hat{e}_{(\mu)}$. We will see that we can also build dual vectors out of vectors, and A^μ and A_μ will be related.

An example of a dual vector is a gradient. Observe

$$\partial_{\alpha'} \phi \equiv \frac{\partial \phi}{\partial x^{\alpha'}} = \frac{\partial x^\beta}{\partial x^{\alpha'}} \frac{\partial \phi}{\partial x^\beta}, \tag{97}$$

simply by the chain rule. We see that $\partial_\mu \phi$ transforms like a dual vector. In components,

$$\partial_\mu \phi = \left(\frac{\partial \phi}{\partial x^0}, \boldsymbol{\nabla} \phi \right). \tag{98}$$

We can go crazy with indices! A ***Cartesian tensor of rank*** (k, ℓ) is an object

$$T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_\ell}. \tag{99}$$

For example a vector is a (1, 0) tensor and a one-form is a (0, 1) tensor. The transformation law is trivial: each index transforms separately. All upper indices transform with a Lorentz transformation, all lower indices transform with an inverse Lorentz transformation

$$T^{\mu_1' \cdots \mu_k'}{}_{\nu_1' \cdots \nu_\ell'} = \Lambda^{\mu_1'}{}_{\mu_1} \Lambda^{\mu_2'}{}_{\mu_2} \cdots \Lambda^{\mu_k'}{}_{\mu_k} \Lambda^{\nu_1}{}_{\nu_1'} \Lambda^{\nu_2}{}_{\nu_2'} \cdots \Lambda^{\nu_\ell}{}_{\nu_\ell'} T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_\ell}. \tag{100}$$

For example,

$$F^{\mu'\nu'} = \Lambda^{\mu'}{}_\alpha \Lambda^{\nu'}{}_\beta F^{\alpha\beta} \tag{101}$$

$$G_{\mu'\nu'} = \Lambda^\alpha{}_{\mu'} \Lambda^\beta{}_{\nu'} G_{\alpha\beta}. \tag{102}$$

Contractions

We have mentioned the summation of upper and lower indices. This is a process called ***contraction***, which we discuss in a little more detail now. Consider contracting a vector and dual vector index:

$$V \cdot A = V^\mu A_\mu = V_\mu A^\mu. \tag{103}$$

The power of this is that it is a scalar. We can easily prove this with our new notation

$$V' \cdot A' = \left(\frac{\partial x^\nu}{\partial x^{\mu'}} V_\nu\right)\left(\frac{\partial x^{\mu'}}{\partial x^\alpha} A^\alpha\right) = \frac{\partial x^\nu}{\partial x^{\mu'}}\frac{\partial x^{\mu'}}{\partial x^\alpha} V_\nu A^\alpha. \tag{104}$$

By the chain rule,

$$\frac{\partial x^\nu}{\partial x^{\mu'}}\frac{\partial x^{\mu'}}{\partial x^\alpha} = \frac{\partial x^\nu}{\partial x^\alpha} = \delta^\nu{}_\alpha, \tag{105}$$

so

$$V' \cdot A' = \delta^\nu{}_\alpha V_\nu A^\alpha = V_\alpha A^\alpha, \tag{106}$$

which proves its a scalar. Dual vectors are good for turning vectors into scalars; they grab a vector and spit out a scalar.

This works in general for any kind of tensors. For example, $T^{\mu\nu} A_\nu = V^\mu$ transforms like a vector, while $T^{\mu\nu} B_{\mu\nu\alpha} = C_\alpha$ is a dual vector. The point is that we can ignore the contracted indices for the purposes of changing frames.

Special Relativity

We have developed a purely mathematical formalism of tensors and how they change under coordinate transformation. Let us now reintroduce special relativity. We define an invariant interval

$$(ds)^2 = \left(dx^0\right)^2 - \left(dx^1\right)^2 - \left(dx^2\right)^2 - \left(dx^3\right)^2 \equiv g_{\alpha\beta} dx^\alpha dx^\beta, \tag{107}$$

where $g_{\alpha\beta}$ is called the ***metric tensor***. In the above coordinates, $g_{00} = 1$, and $g_{ii} = -1$ for $i = 0, \dots, 3$. But this is special to these Cartesian coordinates. Note that $g_{\alpha\beta} = g_{\beta\alpha}$; this is not specific to the above coordinates and will always be true. We write $g_{\alpha\beta} = \eta_{\alpha\beta}$ for the special case of the Minkowski metric above. note that $g_{\mu\nu}$ is a (0, 2) tensor, so an object like $g_{\mu\nu} W^\mu W^\nu$, or of course $g_{\mu\nu} dx^\mu dx^\nu$, is a scalar.

We can relate vectors and dual vectors with the metric by raising and lowering indices. We define

$$V_\mu = g_{\mu\nu} V^\nu, \tag{108}$$

which is a dual vector. To raise indices, we use the inverse metric $g^{\mu\nu}$, which satisfies

$$g_{\mu\nu} g^{\nu\alpha} = \delta_\mu{}^\alpha. \tag{109}$$

For the Minkowski metric η, the inverse metric has the same components as the metric, but this is not true in general. With the inverse metric in hand, a raising operation looks like

$$L^{\rho} = g^{\rho\mu} L_{\mu}. \tag{110}$$

If we have two vectors $V^{\mu} = (V^0, V^1, V^2, V^3)$ and $W^{\mu} = (W^0, W^1, W^2, W^3)$, and we lower the index on W with the Minkowski metric, we get $W_{\mu} = (W^0, -W^1, -W^2, -W^3)$. Thus,

$$V^{\mu} W_{\mu} = V^0 W^0 - V^1 W^1 - V^2 W^2 - V^3 W^3, \tag{111}$$

which is the familiar 4-vector product.

Now consider a derivative

$$\frac{\partial}{\partial x^{\alpha}} = \partial_{\alpha} = \left(\frac{\partial}{\partial x^0}, +\boldsymbol{\nabla} \right). \tag{112}$$

We have seen that this transforms as a dual vector. Raising the index yields

$$\partial^{\alpha} = \frac{\partial}{\partial x_{\alpha}} = \left(\frac{\partial}{\partial x^0}, -\boldsymbol{\nabla} \right). \tag{113}$$

One way to remember where the signs go is to remember that

$$\partial_{\alpha} A^{\alpha} = \frac{\partial A^0}{\partial x^0} + \boldsymbol{\nabla} \cdot \mathbf{A} \tag{114}$$

is an invariant. This looks like a lot of equations we've met in electrodynamics, e.g.

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} = 0. \tag{115}$$

Another important invariant operator is the d'Alembertian

$$\Box = \partial_{\alpha} \partial^{\alpha} = \frac{\partial^2}{\partial (x^0)^2} - \nabla^2 \tag{116}$$

Matrix Representation of Lorentz Group

A useful idea is to write two-index objects as matrices and one-index objects as vectors. For example

$$x^{\mu} \to x = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}, \quad g_{\alpha\beta} \to g = \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & -1 & \\ & & & -1 \end{pmatrix}, \quad x_{\mu} = g_{\mu\nu} x^{\nu} = g \cdot x = \begin{pmatrix} x^0 \\ -x^1 \\ -x^2 \\ -x^3 \end{pmatrix}. \tag{117}$$

A 4-vector product can be written as

$$a \cdot b = (a, gb) = (ga, b) = a^T g b \tag{118}$$

This notation can be very nice when we need to compute, but for abstract manipulations the index notation can be nice too since everything commutes.

The Lorentz group forms the most general set of transformations that preserve the dot product of two 4-vectors. Consider such a transformation $x' = \Lambda x$, or in index notation $x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}$. It obeys

$$(x', gx') = (x, gx) \qquad g_{\mu\nu} x'^{\mu} x'^{\nu} = g_{\rho\sigma} x^{\rho} x^{\sigma} \tag{119}$$

$$x^T \Lambda^T g \Lambda x = x^T g x \qquad g_{\mu\nu} \Lambda^{\mu}{}_{\rho} \Lambda^{\nu}{}_{\sigma} x^{\rho} x^{\sigma} = g_{\rho\sigma} x^{\rho} x^{\sigma} \tag{120}$$

$$\Rightarrow \quad \Lambda^T g \Lambda = g \qquad \Rightarrow \quad g_{\mu\nu} \Lambda^{\mu}{}_{\rho} \Lambda^{\nu}{}_{\sigma} = g_{\rho\sigma} \tag{121}$$

There are two types of Lorentz transformations Λ: ***proper***, in which Λ has an infinitesimal limit and ***improper***, in which it doesn't. Consider the equation $\Lambda^T g \Lambda = g$ derived above. Taking the determinant of both sides, we have

$$\det g = \det\left(\Lambda^T g \Lambda\right) = (\det g)(\det \Lambda)^2, \tag{122}$$

so $\det \Lambda = \pm 1$ for any Lorentz transformation. If Λ has an infinitesimal limit, then we can write it as $\Lambda = 1 + \varepsilon$ for infinitesimal ε, so we must have $\det \Lambda = 1$. Thus, proper Lorentz transformations have $\det \Lambda = 1$. For example, a pure boost along the x-axis is

$$\Lambda = \begin{pmatrix} \gamma & 0 & 0 & \beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta\gamma & 0 & 0 & \gamma \end{pmatrix} \tag{123}$$

As $v \to 0$, $\gamma = 1 + O(v^2/c^2)$ and $\beta\gamma = O(v/c)$, so this matrix has an infinitesimal limit—it is proper. It is simple to show that $\det \Lambda = 1$. Examples of improper transformations include spatial inversions $P = \operatorname{diag}(1, -1, -1, -1)$ and time reversal $T = \operatorname{diag}(-1, 1, 1, 1)$. These both have determinant -1.

The proper transformations are the more interesting ones, so let's consider $\Lambda = 1 + \varepsilon$ or, with indices, $\Lambda^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} + \varepsilon^{\mu}{}_{\nu}$. Plugging this into the defining equation $g_{\mu\nu} \Lambda^{\mu}{}_{\rho} \Lambda^{\nu}{}_{\sigma} = g_{\rho\sigma}$ yields

$$g_{\rho\sigma} = g_{\mu\nu}\left(\delta^{\mu}{}_{\rho} + \varepsilon^{\mu}{}_{\rho}\right)\left(\delta^{\nu}{}_{\sigma} + \varepsilon^{\nu}{}_{\sigma}\right) \tag{124}$$

$$= g_{\rho\sigma} + g_{\mu\nu} \varepsilon^{\mu}{}_{\rho} \delta^{\nu}{}_{\sigma} + g_{\mu\nu} \delta^{\mu}{}_{\rho} \varepsilon^{\nu}{}_{\sigma} \tag{125}$$

$$= g_{\rho\sigma} + \varepsilon_{\sigma\rho} + \varepsilon_{\rho\sigma} \tag{126}$$

In order for this to hold, we need $\varepsilon_{\sigma\rho} = -\varepsilon_{\rho\sigma}$. In principle, $\varepsilon_{\mu\nu}$ could have had 16 components, but antisymmetry implies there are only six independent components. These six components correspond to three boost directions and three rotational directions.

Let's continue to push on this. The coordinates transform as

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} = \left(\delta^{\mu}{}_{\nu} + \varepsilon^{\mu}{}_{\nu}\right) x^{\nu} = x^{\mu} + \delta x^{\mu}, \tag{127}$$

where $\delta x^\mu = \varepsilon^\mu{}_\nu x^\nu$ and $\varepsilon_{\mu\nu} = -\varepsilon_{\nu\mu}$. I claim we can rewrite this in a slightly odd looking way:

$$\delta x^\mu = \frac{i}{2}\varepsilon^{\rho\sigma} L_{\rho\sigma} x^\mu, \tag{128}$$

where

$$L_{\mu\nu} = i[x_\mu \partial_\nu - x_\nu \partial_\mu]. \tag{129}$$

This is like a generalized angular momentum operator. Indeed it would be angular momentum if μ and ν were spatial indices. In (128), the $\varepsilon^{\rho\sigma}$ should be thought of as parameters and the $L_{\rho\sigma}$ as differential operators. Now let's check that (128) is true:

$$\begin{aligned}
\delta x^\mu &= \frac{i}{2}\varepsilon^{\rho\sigma} L_{\rho\sigma} x^\mu & (130)\\
&= -\frac{1}{2}\varepsilon^{\rho\sigma}[x_\rho \partial_\sigma - x_\sigma \partial_\rho] x^\mu & (131)\\
&= -\frac{1}{2}\varepsilon^{\rho\sigma}\left[x_\rho \delta^\mu{}_\sigma - x_\sigma \delta^\mu{}_\rho\right] & (132)\\
&= -\frac{1}{2}\varepsilon^{\rho\mu} x_\rho + \frac{1}{2}\varepsilon^{\mu\sigma} x_\sigma & (133)\\
&= \varepsilon^{\mu\nu} x_\nu = \varepsilon^\mu{}_\nu x^\nu, & (134)
\end{aligned}$$

as desired.

The $L_{\mu\nu}$'s are operators that boost or rotate. Just like angular momentum in quantum mechanics, they have an algebra that can be useful. We compute this below

$$[L_{\mu\nu}, L_{\rho\sigma}] = i g_{\nu\rho} L_{\mu\sigma} - i g_{\mu\rho} L_{\nu\sigma} - i g_{\nu\sigma} L_{\mu\rho} + i g_{\mu\sigma} L_{\nu\rho}. \tag{135}$$

This is a ***Lie algebra***, and the L's are called the ***generators***. Some of these generators are familiar; if we define $L_i \equiv \frac{1}{2}\varepsilon_{ijk} L_{jk}$, then

$$L_1 = \frac{1}{2}(L_{23} - L_{32}) = L_{23}, \qquad L_2 = \frac{1}{2}(L_{31} - L_{13}) = L_{31}, \qquad L_3 = \frac{1}{2}(L_{12} - L_{21}) = L_{12}. \tag{136}$$

We can compute the commutation relations among these operators, for example,

$$[L_1, L_2] = [L_{23}, L_{31}] = i g_{33} L_{21} = -i L_{21} = i L_{12} = i L_3 \tag{137}$$

We find the commutation relations of the angular momentum operators!

These L operators don't know about spin. But we can imagine that we have a particle with internal degrees of freedom that get scrambled under Lorentz transformation. We could have $L_{\mu\nu} + S_{\mu\nu} = M_{\mu\nu}$, where $S_{\mu\nu}$ acts on internal degrees of freedom, has the same commutation relations as $L_{\mu\nu}$, and commutes with $L_{\mu\nu}$. The full Lorentz transformation, acting on spacetime and internal degrees of freedom, is then implemented by $M_{\mu\nu}$.

By defining a few operators, the algebra of the $M_{\mu\nu}$ will become more transparent than what we can see from (135). We define

$$J_i = \frac{1}{2}\varepsilon_{ijk}M_{jk}, \tag{138}$$

where Latin indices take values $1, 2, 3$. We also define

$$K_i = M_{0i}. \tag{139}$$

The three J_i are the generators of rotations and three K_i generate boosts. The commutation relations are

$$[J_i, J_j] = i\varepsilon_{ijk}J_k, \tag{140}$$

so the J_i generate rotations, as we have seen. Commutation relations involving K_i are

$$[J_i, K_j] = i\varepsilon_{ijk}K_k \tag{141}$$

$$[K_i, K_j] = -i\varepsilon_{ijk}J_k. \tag{142}$$

In other words, doing a boost and then a rotation gives us something different from doing the rotation and then the boost. Similarly, noncolinear boosts do not commute. We'll touch on this point a little more in a moment.

Finally, we define

$$A_i = \frac{1}{2}(J_i + iK_i) \quad \text{and} \quad B_i = \frac{1}{2}(J_i - iK_i). \tag{143}$$

We find commutation relations

$$[A_i, B_j] = 0 \tag{144}$$

$$[A_i, A_j] = i\varepsilon_{ijk}A_k \tag{145}$$

$$[B_i, B_j] = i\varepsilon_{ijk}B_k. \tag{146}$$

We see that the Lorentz algebra has been pulled apart into two copies of a simpler algebras, namely the algebra of SU(2). For a single copy of SU(2), such as with angular momentum (or spin), the states are labelled by 2 quantum numbers: j, which tells us the total spin, and m, which for given j takes values $-j \leq m \leq j$, labeling a total of $2j+1$ states. The Lorentz group is described by two pairs of these quantum numbers. Let a and b be the total spin quantum numbers for each pair. Each state is labeled by a, b, m_a, and m_b, i.e.

$$A^2|\psi_{ab}\rangle = a(a+1)|\psi_{ab}\rangle \tag{147}$$

$$B^2|\psi_{ab}\rangle = b(b+1)|\psi_{ab}\rangle \tag{148}$$

From the quantum mechanics of spin, we know the possible values of a and b are $a, b = 0, 1/2, 1, 3/2, \ldots$. The take-away point is that spin comes from the internal transformation properties of particles under the Lorentz algebra.

Under a parity transformation, J_i is an axial vector $J_i \to J_i$ and K_i is a vector $K_i \to -K_i$. This implies that parity switches the A_i and B_i operators. Thus, irreducible representations of the Lorentz group are parity eigenstates if $a = b$. This is not always so, for example the neutrino is left-handed; it has $a = 1/2$ and $b = 0$, so it is not a parity eigenstate.

A scalar particle has spin zero: $a = b = 0$. In contrast, a Dirac particle is a is a linear combination $(a, b) = (1/2, 0) + (0, 1/2)$. We can write a Dirac particle like a column vector

$$\begin{pmatrix} 1 \\ 2 \\ 1' \\ 2' \end{pmatrix}. \tag{149}$$

This is why the Dirac equation has four components.

How do these states transform? We can write an infinitesimal transformation in several different ways

$$D(\varepsilon) = 1 + \frac{i}{2}\varepsilon^{\mu\nu} M_{\mu\nu} \tag{150}$$

$$= 1 + i(\boldsymbol{\theta}_A \cdot \mathbf{A}) + i(\boldsymbol{\theta}_B \cdot \mathbf{B}) \tag{151}$$

$$= 1 + \frac{i}{2}\mathbf{J} \cdot \boldsymbol{\omega} - \mathbf{K} \cdot \boldsymbol{\zeta}. \tag{152}$$

This can be exponentiated in any of the above forms to get the full transformation.

For pure boosts $\boldsymbol{\omega} = 0$ and $\boldsymbol{\zeta} \neq 0$, or equivalently, we choose $\varepsilon^{\mu\nu}$ so that only M_{0i} appears in $D(\varepsilon)$. Choosing $\varepsilon^{10} = -\varepsilon^{01} = \zeta$ to be the only nonzero entries and recalling that $\delta x^\mu = \varepsilon^{\mu\nu} x_\nu$, we find

$$\delta x^0 = \varepsilon^{01} x_1 = -\varepsilon^{10} x_1 = \varepsilon^{10} x^1 \tag{153}$$

$$\delta x^1 = \varepsilon^{10} x_0 = \varepsilon^{10} x^0 \tag{154}$$

In matrix form,

$$\delta \begin{pmatrix} x^0 \\ x^1 \end{pmatrix} = \varepsilon^{10} \begin{pmatrix} x^1 \\ x^0 \end{pmatrix} = \varepsilon^{10} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \end{pmatrix} = \zeta \sigma_x \begin{pmatrix} x^0 \\ x^1 \end{pmatrix}. \tag{155}$$

The transformation exponentiates to

$$D(\zeta) = 1 + \zeta \sigma_x \quad \longrightarrow \quad e^{\zeta \sigma_x} = \begin{pmatrix} \cosh\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix}. \tag{156}$$

So a vector transforms as

$$V' = \begin{pmatrix} \cosh\zeta & \sinh\zeta & 0 & 0 \\ \sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} V, \tag{157}$$

for a boost along the x direction, which of course we already knew. But this provides a more general and abstract framework for doing any kind of Lorentz transformation.

5.4 Thomas Precession

The electron has an intrinsic spin and a related intrinsic magnetic moment. An electron moving around in an atom sees a magnetic field since an electric field in a moving frame becomes a magnetic field. The fine structure of an atom comes from the resulting Zeeman splitting. A relativistic calculation of the fine structure corrects a nonrelativistic calculation by a full factor of 2, not even something $O(v/c)$! The root of the issue is the noncommutation of boost operators.

Imagine a classical (i.e. with well-defined trajectory) spin-1/2 electron orbiting an atomic nucleus. The electron has a total magnetic moment $\boldsymbol{\mu} = \boldsymbol{\mu}_L + \boldsymbol{\mu}_S$, with the first term coming from the orbital motion and the second from its spin. These are given by

$$\boldsymbol{\mu}_L = -\frac{e}{2mc}\mathbf{L} \tag{158}$$

$$\boldsymbol{\mu}_s = -\frac{ge}{2mc}\mathbf{S}, \tag{159}$$

where $g = 2$ very nearly and in our convention $e > 0$. In the lab frame, the electron will feel the electric field from the nucleus, and we could impose our own magnetic field if we wished, but moving to the instantaneous rest frame of the electron will cause these fields to transform into each other. An electron in an atom moves fairly nonrelativistically, in which case the magnetic field transforms as

$$\mathbf{B}' = \mathbf{B} - \frac{\mathbf{v}}{c} \times \mathbf{E} + O\left(\frac{v^2}{c^2}\right). \tag{160}$$

The interaction of $\boldsymbol{\mu}_L$ with the magnetic field in the lab frame gives the normal Zeeman effect, while $\boldsymbol{\mu}_s$ gives rise to the ***anomalous Zeeman effect***. We'll focus on the spin effects. The magnetic field $\mathbf{B}'$ in the instantaneous rest frame gives the electron a potential energy

$$U' = -\boldsymbol{\mu}_S \cdot \mathbf{B}' = \frac{ge}{2mc}\mathbf{S} \cdot \left(\mathbf{B} - \frac{\mathbf{v}}{c} \times \mathbf{E}\right) \tag{161}$$

The electric field in the lab frame $\mathbf{E}$ can be obtained from the Coulomb potential $V = -Ze^2/r$ as

$$-e\mathbf{E} = -\frac{\mathbf{r}}{r}\frac{\partial V}{\partial r} = -\frac{Ze^2\mathbf{r}}{r^3}. \tag{162}$$

Plugging this into U' gives

$$U' = \frac{ge}{2mc}\mathbf{S} \cdot \mathbf{B} - \frac{Zge^2}{2mc^2}\frac{1}{r^3}\mathbf{S} \cdot (\mathbf{v} \times \mathbf{r}) = \frac{ge}{2mc}\mathbf{S} \cdot \mathbf{B} + \frac{Zge^2}{2m^2c^2}\frac{1}{r^3}\mathbf{S} \cdot \mathbf{L}, \tag{163}$$

where we've substituted $\mathbf{L} = \mathbf{r} \times m\mathbf{v}$. The first term gives the correct result for the anomalous Zeeman effect, but the fine structure given by the second term is off by a factor of 2. The reason is that we've only boosted to get into the electron's rest frame, we haven't accounted for its rotation in any way. Of course, this is also the energy of the electron in its instantaneous rest frame, and we need the energy in the lab frame.

The energy of a magnetic dipole is derived from its equation of motion

$$\tau = \boldsymbol{\mu} \times \mathbf{B}, \tag{164}$$

where τ is the torque on the dipole. We can find the energy of the electron in the lab frame by casting $\tau = d\mathbf{S}/dt$ as seen in the lab frame in a similar way. Suppose the electron's rest frame rotates at a frequency $\boldsymbol{\omega}_T$, called the ***Thomas frequency***, which we'll derive shortly. From classical mechanics, the rate of change of a vector as seen in a stationary frame is related to its rate of change in a rotating frame by

$$\left.\frac{d\mathbf{G}}{dt}\right|_{\text{stat}} = \left.\frac{d\mathbf{G}}{dt}\right|_{\text{rot}} + \boldsymbol{\omega} \times \mathbf{G}. \tag{165}$$

For the electron, we find

$$\left.\frac{d\mathbf{S}}{dt}\right|_{\text{stat}} = \left.\frac{d\mathbf{S}}{dt}\right|_{\text{rot}} + \boldsymbol{\omega}_T \times \mathbf{S} \tag{166}$$

$$= \boldsymbol{\mu}_s \times \mathbf{B}' + \boldsymbol{\omega}_T \times \mathbf{S} \tag{167}$$

$$= -\frac{ge}{2mc}\mathbf{S} \times \left(\mathbf{B}' + \frac{2mc}{ge}\boldsymbol{\omega}_T\right) \tag{168}$$

This tells us that the energy in the lab frame is

$$U = U' - \frac{2mc}{ge}\boldsymbol{\mu}_s \cdot \boldsymbol{\omega}_T = U' + \mathbf{S} \cdot \boldsymbol{\omega}_T. \tag{169}$$

We shall see that the $\mathbf{S} \cdot \boldsymbol{\omega}_T$ term accurately corrects the factor of two we missed before.

We now endeavor to derive $\boldsymbol{\omega}_T$. Suppose that at time t, the velocity of the electron is $\mathbf{v}(t) = c\boldsymbol{\beta}$. Transforming to the instantaneous rest frame changes the spacetime coordinates to

$$x' = A(\boldsymbol{\beta})x, \tag{170}$$

where $A(\boldsymbol{\beta})$ is the boost matrix for given $\boldsymbol{\beta}$. At a later time $\mathbf{v}(t+\delta t) = c(\boldsymbol{\beta} + \delta\boldsymbol{\beta})$, so to get to the electron's rest frame at $t + \delta t$ we need a new transformation

$$x'' = A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})x \tag{171}$$

How do we transform from x' to x''? In other words, what matrix B satisfies $x'' = Bx'$? We can go from the electron's rest frame at x' to the lab frame, and then back to the electron's rest frame at x'', so

$$B = A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})A^{-1}(\boldsymbol{\beta}) = A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})A(-\boldsymbol{\beta}). \tag{172}$$

If the electron is moving around the nucleus, then these boosts are not in the same direction. The product will be equal to a pure boost *and* a pure rotation

$$A(\boldsymbol{\beta} + \delta\boldsymbol{\beta})A(-\boldsymbol{\beta}) \simeq R(\Delta\boldsymbol{\Omega})A(\Delta\boldsymbol{\beta}) \tag{173}$$

The pure boost gives us $\mathbf{B}' = \mathbf{B} - (\mathbf{v}/c) \times \mathbf{E}$, but the rotation is what gives us the Thomas precession by frequency $\boldsymbol{\omega}_T$.

We want to find $\Delta\Omega$, since then we can get the Thomas frequency by $\boldsymbol{\omega}_T = \Delta\Omega/\delta t$. We could work with four by four matrices and work out the transformation, which is how Jackson does it. But there is a more elegant way using the formalism we have developed. The full transformation can be written in exponential form as

$$B = \exp\left[-\boldsymbol{\zeta}(\boldsymbol{\beta} + \delta\boldsymbol{\beta}) \cdot \mathbf{K}\right] \exp\left[\boldsymbol{\zeta}(\boldsymbol{\beta}) \cdot \mathbf{K}\right], \tag{174}$$

where $\boldsymbol{\zeta}(\boldsymbol{\beta})$ is the rapidity associated with $\boldsymbol{\beta}$ and the K_i are the 4×4 matrices that generate boosts. We compare this to a single boost without rotation

$$B' = \exp\left[(-\boldsymbol{\zeta}(\boldsymbol{\beta} + \delta\boldsymbol{\beta}) + \boldsymbol{\zeta}(\boldsymbol{\beta})) \cdot \mathbf{K}\right] \tag{175}$$

For ease of notation, let us define

$$\boldsymbol{\zeta} = \boldsymbol{\zeta}(\boldsymbol{\beta} + \delta\boldsymbol{\beta}) \tag{176}$$

$$\boldsymbol{\zeta}' = \boldsymbol{\zeta}(\boldsymbol{\beta}). \tag{177}$$

In the nonrelativistic limit, the rapidity is approximately $\boldsymbol{\zeta}(\boldsymbol{\beta}) \simeq \boldsymbol{\beta} \ll 1$, so we can expand the exponentials! Being very careful not to commute any K's, we expand

$$B = \left[1 - \boldsymbol{\zeta} \cdot \mathbf{K} + \frac{1}{2}(\boldsymbol{\zeta} \cdot \mathbf{K})^2 + \cdots\right]\left[1 + \boldsymbol{\zeta}' \cdot \mathbf{K} + \frac{1}{2}(\boldsymbol{\zeta}' \cdot \mathbf{K})^2 + \cdots\right] \tag{178}$$

$$= 1 + (\boldsymbol{\zeta}' - \boldsymbol{\zeta}) \cdot \mathbf{K} + K_i K_j \left(\frac{1}{2}\zeta_i\zeta_j - \zeta_i\zeta_j' + \frac{1}{2}\zeta_i'\zeta_j'\right) \tag{179}$$

In contrast, we have

$$B' = A(\Delta\boldsymbol{\beta}) = 1 + (\boldsymbol{\zeta}' - \boldsymbol{\zeta}) \cdot \mathbf{K} + \frac{1}{2}K_i K_j (\zeta' - \zeta)_i (\zeta' - \zeta)_j. \tag{180}$$

We notice that $B \neq B'$ at second order. Explicitly, the difference is

$$B' - B = \frac{1}{2}K_i K_j \left\{\zeta_i\zeta_j' - \zeta_i'\zeta_j\right\} \tag{181}$$

$$= \frac{1}{2}\zeta_i\zeta_j'[K_i K_j - K_j K_i] \tag{182}$$

$$= \frac{1}{2}\zeta_i\zeta_j'[K_i, K_j]. \tag{183}$$

The commutator of boosts in different directions is a rotation operator $[K_i, K_j] = -i\varepsilon_{ijk}J_k$. Here we're interested in the spin interactions so we'll take $[K_i, K_j] = -i\varepsilon_{ijk}S_k$. Thus, to second order in ζ, ζ', we have

$$B = B' + \frac{i}{2}\zeta_i\zeta_j'\varepsilon_{ijk}S_k = B' + \frac{i}{2}\mathbf{S} \cdot (\boldsymbol{\zeta} \times \boldsymbol{\zeta}') = \left[1 + \frac{i}{2}\mathbf{S} \cdot (\boldsymbol{\zeta} \times \boldsymbol{\zeta}')\right] B' \tag{184}$$

From this and the relation $B = R(\Delta\boldsymbol{\Omega})B'$ we can read off the rotation angle $\Delta\boldsymbol{\Omega} = \frac{1}{2}\boldsymbol{\zeta} \times \boldsymbol{\zeta}'$. For small velocity, $\boldsymbol{\zeta}(\boldsymbol{\beta}) \simeq \boldsymbol{\beta}$, so

$$\Delta\boldsymbol{\Omega} = -\frac{1}{2}\boldsymbol{\beta} \times \delta\boldsymbol{\beta} \tag{185}$$

In a fully relativistic treatment the 1/2 is a $\gamma^2/(1+\gamma^2)$, but we're being nonrelativistic. Thus, the Thomas frequency is

$$\boldsymbol{\omega}_T = \frac{\Delta\boldsymbol{\Omega}}{\delta t} = -\frac{1}{2}\boldsymbol{\beta} \times \frac{\delta\boldsymbol{\beta}}{\delta t} = \frac{1}{2c^2}\mathbf{a} \times \mathbf{v}. \tag{186}$$

For the electron in the Coulomb potential,

$$\mathbf{a} = \frac{\mathbf{F}}{m} = -\frac{\mathbf{r}}{mr}\frac{\partial V}{\partial r} = -\frac{Ze^2}{mr^3}\mathbf{r}, \tag{187}$$

so the Thomas frequency is

$$\boldsymbol{\omega}_T = -\frac{Ze^2}{2mc^2r^3}\mathbf{r} \times \mathbf{v} = -\frac{Ze^2}{2m^2c^2r^3}\mathbf{L}. \tag{188}$$

Finally, plugging this into the interaction energy yields

$$U = \frac{ge}{2mc}\mathbf{S} \cdot \mathbf{B} + \frac{Ze^2(g-1)}{2m^2c^2r^3}\mathbf{S} \cdot \mathbf{L}, \tag{189}$$

with the -1 in the $g-1$ factor coming from the Thomas precession. Note that the Thomas precession includes no factor of g since it is purely geometrical. Since $g \simeq 2$, we see that $g-1$ is different from g by a factor of 2, so the above expression gives the correct fine structure. Hooray!

6 Classical Field Theory

Our goal now is to construct a formalism that will give us equations of motion for fields and particles while maintaining manifest Lorentz invariance. To get a sense of the kind of questions we should be asking, let's start by looking at the continuity equation

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{J} = 0. \tag{190}$$

If we write $J^\mu = (c\rho, \mathbf{J})$, then we can write this as

$$\partial_\mu J^\mu = 0. \tag{191}$$

This is invariant, so if we have charge conservation in one frame then we have it in all frames.

We should ask whether it is sensible for J^μ to be a four-vector or not. We know the charge

$$Q = \int \rho \, d^3x \tag{192}$$

is conserved. Is the charge the same in all frames? This is tricky since d^3x is not Lorentz invariant. However, d^4x is Lorentz invariant, and we can write $d^4x = dx^0\, d^3x$. We see that if Q is the same in all frames, then ρ must transform like dx^0.

The Lorentz gauge condition is

$$\frac{1}{c}\frac{\partial \Phi}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{A} = 0 \tag{193}$$

The wave equations were

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2\right)\begin{pmatrix}\Phi\\ \mathbf{A}\end{pmatrix} = \partial_\mu \partial^\mu \begin{pmatrix}\Phi\\ \mathbf{A}\end{pmatrix} = \frac{4\pi}{c}\begin{pmatrix}c\rho\\ \mathbf{J}\end{pmatrix}.$$

We know $\partial_\mu \partial^\mu$ is invariant, so if $(c\rho, \mathbf{J})$ is a four-vector then so is $A^\mu = (\Phi, \mathbf{A})$. We can write

$$\Box A^\mu = \frac{4\pi}{c} J^\mu,$$

where $\Box = \partial_\mu \partial^\mu$.

But we should be worried about gauge choices. What if we wanted to work in Coulomb gauge $\boldsymbol{\nabla} \cdot \mathbf{A} = 0$. This is not a covariant gauge; this equation doesn't survive a Lorentz transformation. Would the predictions of special relativity still be consistent with electrodynamics if we worked in this gauge? They are, but it is not necessarily easy to show.

Instead of starting with the equations of motion and trying to prove Lorentz invariance and gauge invariance, we're going to go the other way. We ask what theoretical description of nature is consistent with the symmetries of Lorentz invariance and our gauge transformations. We'll construct a classical Lagrangian for electromagnetism determined by the constraints imposed by these symmetries. This is the modern way to do everything. The key is that we give priority to the symmetries. This allows us to ask "why" questions, and determine the degree of uniqueness of our equations of motion. Furthermore, if we build relativistic invariance into our Lagrangian from the start, then we know that what we derive will be relativistically invariant. Having a Lagrangian also gives us access to a Hamiltonian, which allows us to move into the quantum theory.

In classical particle mechanics, we postulate an action

$$S = \int dt\, L\left(q_i, \frac{dq_i}{dt}\right).$$

We get equations of motion by forcing $\delta S = 0$ under variations of the coordinates q_i, dq_i/dt. The mechanics for classical fields is almost the same thing. We begin with the field variables $\varphi(x,t)$, $\partial\varphi/\partial x^j$, and $\partial\varphi/\partial t$. We could in general have many fields, in which case we would have to include an index like φ_j. The analog of the Lagrangian is the Lagrange density

$$L = \int d^3x\, \mathcal{L}(\varphi_j, \text{derivatives}),$$

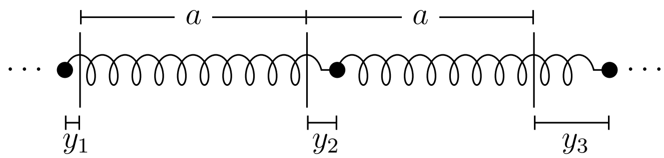
and we can substitute this into the action to obtain

$$S = \int d^4x \, \mathcal{L}(\varphi_j, \partial_\mu \varphi_j).$$

We could in principle have included higher order derivatives, but let's not get too crazy. The Euler Lagrange equations become

$$\partial_\mu \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi_j)} \right] - \frac{\partial \mathcal{L}}{\partial \varphi_j} = 0.$$

Example 2. Imagine we have coupled harmonic oscillators which, in equilibrium, are an equal distance a apart. We define $y_j = y(x_j, t)$ to be the displacement of the jth mass from equilibrium.



The Lagrangian is

$$L = \sum_{j=1}^{N} \frac{1}{2} m \dot{y}_j^2 - \frac{1}{2} k (y_j - y_{j+1})^2 - \tilde{V}(y_j),$$

for some potential $\tilde{V}$. The equations of motion are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{y}_j} - \frac{\partial L}{\partial y_j} = m\ddot{y}_j - k(y_{j-1} - 2y_j + y_{j+1}) + \frac{\partial \tilde{V}}{\partial y_j} = 0.$$

To get to a classical field theory, we take $a \to 0$ and suppose that the y vary smoothly with x, i.e.

$$y_{j+1} = y_j + a \left.\frac{dy}{dx}\right|_{x=x_j} + \frac{1}{2} a^2 \left.\frac{d^2 y}{dx^2}\right|_{x=x_j}$$

The equation of motion is

$$m\ddot{y}(x) - ka^2 \frac{d^2 y}{dx^2} + \frac{\partial \tilde{V}}{\partial y(x)} = 0.$$

We define $\varphi(x,t) = y(x,t)/ka^2$, $1/c^2 = m/ka^2$, and $V = \tilde{V}/(ka^2)^3$. Then

$$\frac{1}{c^2}\frac{\partial^2 \varphi}{\partial t^2} - \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial V}{\partial \varphi} = \partial_\mu \partial^\mu \varphi + \frac{\partial V}{\partial \varphi} = 0.$$

This is the equation of motion obtained from a Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \varphi)(\partial^\mu \varphi) - V(\varphi),$$

which describes the continuum limit of our discrete Lagrangian.

6.1 Noether's Theorem

Suppose you had a classical particle Lagrangian that was only a function of $\dot{q}$, i.e. $L = L(\dot{q})$. Then the Euler-Lagrange equations reduce to

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0. \tag{194}$$

Thus, $\partial L/\partial \dot{q}$ is a conserved quantity. Not having q in our Lagrangian is a statement of translation symmetry. Thus, we have a conserved quantity associated with a symmetry.

Let's go back to the field theory. We make a change $\varphi_j(x,t) \to \varphi_j(x,t) + \delta\varphi_j(x,t)$. The derivatives transform as

$$\partial_\mu \varphi_j \to \partial_\mu \varphi_j(x,t) + \delta\partial_\mu \varphi_j(x,t) = \partial_\mu \varphi_j(x,t) + \partial_\mu \delta\varphi_j(x,t). \tag{195}$$

The Lagrange density $\mathcal{L}$ changes as we make this change. If the change is small, then we can Taylor expand to first order

$$\delta\mathcal{L} = \sum_j \frac{\partial \mathcal{L}}{\partial \varphi_j}\delta\varphi_j + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi_j)}\delta(\partial_\mu \varphi_j). \tag{196}$$

But the two partial derivatives above are not independent due to the Euler-Lagrange equations. Substituting the Euler-Lagrange equations in yields

$$\delta\mathcal{L} = \sum_j \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi_j)}\delta\varphi_j + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi_j)}\delta(\partial_\mu \varphi_j) = \partial_\mu \sum_j \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi_j)}\delta\varphi_j \tag{197}$$

We define the Noether current

$$J^\mu = \sum_j \frac{\partial \mathcal{L}}{\partial(\partial_\mu \varphi_j)}\delta\varphi_j \tag{198}$$

Supposing the Lagrangian is invariant $\delta\mathcal{L} = 0$, then (197) implies a conservation law $\partial_\mu J^\mu = 0$. We call J^μ a ***conserved current***. We see that symmetries, i.e. transformations for which $\delta\mathcal{L} = 0$, imply conservation laws. We can integrate J^0 to get

$$Q = \int d^3x \, J^0. \tag{199}$$

This is a conserved charge and is automatically invariant under change of coordinates.

Changes in φ can be "internal," e.g. $\varphi(x,t) \to \varphi'(x,t)$, as opposed to spacetime transformations $\varphi(x,t) \to \varphi(x',t')$. One possibility is that the symmetry can be "global," in which the transformation law for φ does not depend on x, for example $\varphi(x,t) \to e^{i\theta}\varphi(x,t)$ or $\varphi(x,t) \to \varphi(x,t) + c$ for constant θ and c. Alternatively, we can do "local" internal transformations, such as

$$\varphi(x,t) \to e^{i\theta(x,t)}\varphi(x,t). \tag{200}$$

Generally, the global symmetries tell us about conserved quantities and the local ones tell us about interactions, as we'll see.

Example 3. Consider a classical complex scalar field $\varphi(x,t) = \varphi_1(x,t) + i\varphi_2(x,t)$. We could think of this like the wavefunction for a BEC in a trap, or any "classical" Schrödinger wavefunction. Alternatively, you could think of it as a collection of classical spins in a plane. Consider a symmetry transformation

$$\varphi'(x,t) = e^{i\theta}\varphi(x,t) \quad \Longleftrightarrow \quad \begin{pmatrix} \varphi_1' \\ \varphi_2' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} \tag{201}$$

Let's consider a toy Lagrangian

$$\mathcal{L} = \frac{1}{2}\left[(\partial_\mu\varphi_1)^2 + (\partial_\mu\varphi_2)^2\right] - V(\varphi_1,\varphi_2) = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi^* - V(\varphi,\varphi^*). \tag{202}$$

In the middle expression $(\partial_\mu\varphi_1)^2 = \partial_\mu\varphi_1\partial^\mu\varphi_1$. It just saves a little space to write it this way. You could also write $|\partial_\mu\varphi|^2 = \partial_\mu\varphi\partial^\mu\varphi^*$. Why do we include the 1/2? We can compute

$$\partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi_1)} = \partial_\mu\partial^\mu\varphi_1, \tag{203}$$

which doesn't have a 1/2. So it's just a matter of convention for making our conserved current and continuity equation look nicer.

For an infinitesimal transformation, $\theta \ll 1$, we have

$$\begin{pmatrix} \varphi_1' \\ \varphi_2' \end{pmatrix} = \begin{pmatrix} 1 & -\theta \\ \theta & 1 \end{pmatrix}\begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}. \tag{204}$$

Explicitly, the variations in the fields are then

$$\delta\varphi_1 = -\theta\varphi_2 \tag{205}$$
$$\delta\varphi_2 = \theta\varphi_1 \tag{206}$$
$$\delta(\partial_\mu\varphi_1) = -\theta\partial_\mu\varphi_2 \tag{207}$$
$$\delta(\partial_\mu\varphi_2) = \theta\partial_\mu\varphi_1. \tag{208}$$

Now, these transformations will be a symmetry of the Lagrangian if $V(\varphi_1,\varphi_2) = V(\varphi_1^2+\varphi_2^2)$. Then the potential will clearly be invariant, and will not contribute to $\delta\mathcal{L}$. The variation of the remaining term (the "kinetic" term) is

$$\delta\mathcal{L}_{\text{kin}} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi_1)}\delta(\partial_\mu\varphi_1) + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi_2)}\delta(\partial_\mu\varphi_2) \tag{209}$$
$$= -(\partial^\mu\varphi_1)(\theta\partial_\mu\varphi_2) + (\partial^\mu\varphi_2)(\theta\partial_\mu\varphi_1) = 0. \tag{210}$$

So the transformation is indeed a symmetry. The conserved current is

$$J^\mu = \sum_{j=1}^{2}\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi_j)}\delta\varphi_j \tag{211}$$
$$= -\partial^\mu\varphi_1\theta\varphi_2 + \partial^\mu\varphi_2\theta\varphi_1. \tag{212}$$

Usually, we scale out the θ since it's not physical, and we don't lose anything by removing it. The current is then

$$J^\mu = \varphi_1\partial^\mu\varphi_2 - \varphi_2\partial^\mu\varphi_1. \tag{213}$$

6.2 Gauge Transformations

Let's return to our complex scalar field theory with Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi^* - V(\varphi, \varphi^*), \tag{214}$$

where the potential is a function only of the magnitude $V(\varphi, \varphi^*) = V(|\varphi|)$. So far, we've considered the infinitesimal transformation

$$\delta\varphi = i\theta\varphi \tag{215}$$

$$\delta\varphi^* = -i\theta\varphi^*, \tag{216}$$

or $\varphi \to e^{i\theta}\varphi$, where θ was a constant over spacetime. We saw that there was a conserved current

$$J^\mu = \varphi_1\partial^\mu\varphi_2 - \varphi_2\partial^\mu\varphi_1 = \frac{i}{2}(\varphi\partial^\mu\varphi^* - \varphi^*\partial^\mu\varphi). \tag{217}$$

Now we do something crazy: we let θ vary from point to point in spacetime. What if we want that transformation to still be a symmetry? We'll set $\theta(x,t) = q\varepsilon(x,t)$, where q is constant, and redefine the current by qJ^μ. This q is like the charge of the field (or the particle it represents), so including it in the current gives something with units of charge worked in, which is what we're used to in a current. The variation in the Lagrangian is still

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\varphi}\delta\varphi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}\delta(\partial_\mu\varphi) + (\varphi \to \varphi^*), \tag{218}$$

but the variation in the fields is now

$$\delta\varphi = iq\varepsilon\varphi \tag{219}$$

$$\delta\partial_\mu\varphi = \partial_\mu\delta\varphi = iq\varepsilon\partial_\mu\varphi + iq\varphi\partial_\mu\varepsilon. \tag{220}$$

The variation in φ is essentially the same, but the second term in the variation of $\partial_\mu\varphi$ is new! The equation of motion is still

$$\partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)} = \frac{\partial\mathcal{L}}{\partial\varphi}. \tag{221}$$

Plugging the equation of motion and the variations into (218) yields

$$\delta\mathcal{L} = \varepsilon\partial_\mu J^\mu(x) + \partial_\mu\varepsilon\left[\frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi)}iq\varphi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\varphi^*)}(-iq\varphi^*)\right] \tag{222}$$

The object in brackets has significance: it is the current J^μ! So the variation of the Lagrangian is

$$\delta\mathcal{L} = \epsilon(x,t)\partial_\mu J^\mu + J^\mu\partial_\mu\epsilon(x,t). \tag{223}$$

The first term is zero since the current is conserved. Even if we started with a local transformation, we can always take the local transformation to be global and show that $\partial_\mu J^\mu = 0$ a la Noether's theorem. The second term looks like garbage, but it is coupling to the conserved current, which is interesting. We wanted to set $\delta\mathcal{L} = 0$, but we seem to have failed, since ε is arbitrary and so is $\partial_\mu\varepsilon(x,t)$. The only way to make this a symmetry is to add new degrees of freedom. We want to cancel the $\partial_\mu\varepsilon$ term, which is a four-vector, so we'll add a four-vector field A_μ and give it the transformation

$$\delta A_\mu = \partial_\mu\varepsilon. \tag{224}$$

This is a gauge transformation. It must compensate for the rotation of the φ field at every point. In more familiar notation,

$$\mathbf{A}' = \mathbf{A} - \boldsymbol{\nabla}\varepsilon \tag{225}$$

$$(A^0)' = A^0 + \frac{1}{c}\frac{\partial\varepsilon}{\partial t}, \tag{226}$$

where the minus sign comes from raising the index. Note that (226) is the transformation of the scalar potential. We should therefore identify $A^0 = \Phi$.

Let's see what this gets us. We'll have a new Lagrangian $\mathcal{L}' = \mathcal{L}_\varphi + \mathcal{L}$ that depends on the A_μ and its derivatives, just like it does for φ. Here $\mathcal{L}_\varphi$ is the old Lagrangian and $\mathcal{L}$ is a new piece that includes the A_μ fields. We want $\delta\mathcal{L}' = 0$ with the given transformations. The new change is

$$\delta\mathcal{L}' = \delta\mathcal{L}_\varphi + \frac{\partial\mathcal{L}}{\partial A_\mu}\delta A_\mu + \frac{\partial\mathcal{L}}{\partial(\partial_\nu A_\mu)}\partial_\nu\delta A_\mu. \tag{227}$$

$$= J^\mu\partial_\mu\varepsilon + \frac{\partial\mathcal{L}}{\partial A_\mu}\partial_\mu\varepsilon + \frac{\partial\mathcal{L}}{\partial(\partial_\nu A_\mu)}\partial_\nu\partial_\mu\varepsilon. \tag{228}$$

We can get the first two terms to cancel if

$$\frac{\partial\mathcal{L}}{\partial A_\mu} = -J^\mu. \tag{229}$$

Thus, the new degree of freedom couples to the conserved current in a way that is forced on us by $\delta\mathcal{L}' = 0$. There is a unique specification of the coupling of the new field to the conserved current by the interaction $\mathcal{L}_I = -J_\mu A^\mu$. To kill the second term, we notice that $\partial_\nu\partial_\mu\varepsilon$ is symmetric in ν,μ. Therefore this last term is zero if

$$\frac{\partial\mathcal{L}}{\partial(\partial_\nu A_\lambda)} = -\frac{\partial\mathcal{L}}{\partial(\partial_\lambda A_\nu)}, \tag{230}$$

which tells us that the derivatives of A enter the Lagrangian as

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{231}$$

In summary, the total Lagrangian is

$$\mathcal{L}(\varphi, \partial_\mu\varphi, A_\mu, \partial_\mu A_\nu) = \mathcal{L}(\varphi, \partial_\mu\varphi) - J_\mu A^\mu + \mathcal{L}(F_{\mu\nu}). \tag{232}$$

We call the spacetime dependent transformations ***local gauge transformations*** and we call theories with local gauge transformations ***gauge theories***. Nature is generally described by gauge theories. Generalizations of our example could have φ be a column vector, and the gauge transformation could be $\varphi \to R(x,t)\varphi$, where R is a matrix. In these generalizations, A_μ becomes a matrix as well.

Covariant Derivatives

We'll now work to rewrite our Lagrangian in such a way that the gauge symmetry is manifest. It will take the form

$$\mathcal{L}(\varphi, D_\mu\varphi, F_{\mu\nu}), \tag{233}$$

where $D_\mu\varphi$ is a ***covariant derivative*** that will bundle together the derivative of φ and its interactions with the vector field. We define it as

$$D_\mu\varphi = \partial_\mu\varphi - iqA_\mu\varphi. \tag{234}$$

In general, if we want to "gauge" a global symmetry we can just replace ordinary derivatives by covariant derivatives.

The full gauge transformation is

$$\varphi' = e^{iq\varepsilon(x)}\varphi \tag{235}$$

$$A'_\mu = A_\mu + \partial_\mu\varepsilon \tag{236}$$

The derivatives appeared in the original Lagrangian in the term $\partial_\mu\varphi\partial^\mu\varphi^*$. We then replace ∂_μ by D_μ to get the term $D_\mu\varphi D^\mu\varphi^*$. The idea behind the covariant derivative is that φ and $D_\mu\varphi$ change in the same way under a gauge transformation, or rather $D_\mu\varphi$ changes under gauge transformations like $\partial_\mu\varphi$ changes under global transformations. Let's check that:

$$D'_\mu\varphi' = \partial_\mu\varphi' - iqA'_\mu\varphi' \tag{237}$$

$$= \partial_\mu\big(e^{iq\varepsilon}\varphi\big) - iq(A_\mu + \partial_\mu\varepsilon)e^{iq\varepsilon}\varphi \tag{238}$$

$$= e^{iq\varepsilon}(\partial_\mu\varphi + \cancel{iq\partial_\mu\varepsilon\varphi} - iqA_\mu\varphi - \cancel{iq\partial_\mu\varepsilon\varphi}) \tag{239}$$

$$= e^{iq\varepsilon}(\partial_\mu\varphi - iqA_\mu\varphi) = e^{iq\varepsilon}D_\mu\varphi, \tag{240}$$

which mimics the transformation of φ, as desired. We can therefore construct a gauge invariant Lagrangian

$$\mathcal{L} = \frac{1}{2}(D_\mu\varphi)^*(D^\mu\varphi) - V(|\varphi|), \tag{241}$$

since the transformation of $D^\mu\varphi$ and $(D_\mu\varphi)^*$ will cancel each other. In general we might separate the full Lagrangian into two pieces

$$\mathcal{L} = \mathcal{L}(\varphi, D_\mu\varphi) + \mathcal{L}(F_{\mu\nu}). \tag{242}$$

6.3 Maxwell's Equations

We still need the term $\mathcal{L}(F_{\mu\nu})$, otherwise there would be no dynamics behind the gauge fields. They would just be random variables. What is a good candidate action for $F_{\mu\nu}$? The action

$$S = \int d^4x\, \mathcal{L} \tag{243}$$

must be a scalar. The measure d^4x is invariant, so $\mathcal{L}$ must be a scalar as well. So we could take

$$\mathcal{L} = c_1 F_{\mu\nu}F^{\mu\nu} + c_2(F_{\mu\nu}F^{\mu\nu})^2 + \cdots \tag{244}$$

where $c_1, c_2, \ldots$ are coefficients. In fact, the Lagrangian for electrodynamics in cgs is

$$\mathcal{L} = -\frac{1}{16\pi}F_{\mu\nu}F^{\mu\nu} - \frac{1}{c}J_\mu A^\mu. \tag{245}$$

Why isn't there more? Well, we'll come back to that later.

We get Maxwell's equations out of the above Lagrangian via the Euler-Lagrange equations:

$$0 = \partial^\nu \frac{\partial \mathcal{L}}{\partial(\partial^\nu A^\mu)} - \frac{\partial \mathcal{L}}{\partial A^\mu}. \tag{246}$$

Note that there are four equations here, one for each μ. Since this Lagrangian is a little more complicated, we'll take the derivatives very methodically. The key is to remember that there are hidden metrics when indices are raised or lowered. We write

$$\mathcal{L} = -\frac{1}{16\pi}g_{\lambda\mu}g_{\sigma\nu}\left(\partial^\lambda A^\sigma - \partial^\sigma A^\lambda\right)(\partial^\mu A^\nu - \partial^\nu A^\mu) - \frac{1}{c}J_\alpha A^\alpha. \tag{247}$$

Then derivatives of the Lagrangian follow from

$$\frac{\partial(\partial^\mu A^\nu)}{\partial(\partial^\beta A^\alpha)} = \delta^\mu_\beta \delta^\nu_\alpha. \tag{248}$$

We compute

$$\frac{\partial \mathcal{L}}{\partial(\partial^\beta A^\alpha)} = -\frac{1}{16\pi}g_{\lambda\mu}g_{\sigma\nu}\left(\delta^\lambda_\beta\delta^\sigma_\alpha F^{\mu\nu} - \delta^\sigma_\beta\delta^\lambda_\alpha F^{\mu\nu} + F^{\lambda\sigma}\delta^\mu_\beta\delta^\nu_\alpha - F^{\lambda\sigma}\delta^\nu_\beta\delta^\mu_\alpha\right) \tag{249}$$

$$= -\frac{1}{16\pi}(F_{\beta\alpha} - F_{\alpha\beta} + F_{\beta\alpha} - F_{\alpha\beta}) \tag{250}$$

$$= -\frac{1}{4\pi}F_{\beta\alpha}, \tag{251}$$

where we have used $F_{\alpha\beta} = -F_{\beta\alpha}$ in the last line. The second term in the Euler-Lagrange equations is easy

$$\frac{\partial \mathcal{L}}{\partial A^\alpha} = -\frac{1}{c}J_\alpha. \tag{252}$$

Therefore, our equation of motion is

$$\boxed{\partial^\beta F_{\beta\alpha} = \frac{4\pi}{c} J_\alpha.} \tag{253}$$

This probably doesn't look so familiar. Recall that $A^\mu = (\Phi, \mathbf{A})$, so $A_\mu = (\Phi, -\mathbf{A})$ and that the fields are written in terms of the potentials as

$$B_i = (\boldsymbol{\nabla} \times \mathbf{A})_i = -\varepsilon_{ijk}\partial_j A_k \tag{254}$$

$$E_i = \left(-\boldsymbol{\nabla}\Phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}\right)_i = -\partial_i A_0 + \partial_0 A_i. \tag{255}$$

Recall that $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Therefore, in matrix form we can write

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix} \tag{256}$$

You can check that with the indices raised, it is

$$F^{\alpha\beta} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}. \tag{257}$$

Now we return to the covariant Maxwell equations $\partial^\beta F_{\beta\alpha} = (4\pi/c)J_\alpha$. Recall that $J^\mu = (c\rho, \mathbf{J})$ so $J_\mu = (c\rho, -\mathbf{J})$. Then the first ($\alpha = 0$) covariant Maxwell equation is

$$\boldsymbol{\nabla} \cdot \mathbf{E} = 4\pi\rho. \tag{258}$$

while the spatial ($\alpha = 1, 2, 3$) covariant Maxwell equations are

$$\boldsymbol{\nabla} \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{J}. \tag{259}$$

So we have two of Maxwell's equations.

The other two are a little different. They are immediately satisfied once we define $\mathbf{A}$ and Φ. In other words, they are satisfied simply because $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ by definition. Automatically we have

$$0 = \partial^\alpha F^{\beta\gamma} + \partial^\beta F^{\gamma\alpha} + \partial^\gamma F^{\alpha\beta}, \tag{260}$$

just from the fact that $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. Perhaps a nicer way to write this is to introduce the antisymmetric ***dual field tensor*** $\mathcal{F}^{\alpha\beta}$, defined by

$$\mathcal{F}^{\alpha\beta} = \frac{1}{2}\varepsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta} = \begin{pmatrix} 0 & -B_x & -B_y & -B_z \\ B_x & 0 & E_z & -E_y \\ B_y & -E_z & 0 & E_x \\ B_z & E_y & -E_x & 0 \end{pmatrix}, \tag{261}$$

where $\epsilon^{\alpha\beta\gamma\delta}$ is the totally antisymmetric Levi-Civita tensor we four, four-valued indices. Then we can write the homogeneous Maxwell equations as

$$\partial_\alpha \mathcal{F}^{\alpha\beta} = \frac{1}{2}\varepsilon^{\alpha\beta\mu\nu}\partial_\alpha(\partial_\mu A_\nu - \partial_\nu A_\mu) = 0, \tag{262}$$

where the cancellation is apparent since $\varepsilon^{\alpha\beta\mu\nu}$ is antisymmetric and two partial derivatives $\partial_\alpha\partial_\mu$ are symmetric.

Lorentz Transformation of Electric and Magnetic Fields

Recall that the potential $A^\mu = (\Phi, \mathbf{A})$ is a 4-vector, so that it transforms simply as

$$A'^\mu = \Lambda^\mu{}_\nu A^\nu. \tag{263}$$

On the other hand, the $\mathbf{E}$ and $\mathbf{B}$ fields live in a rank two tensor $F_{\mu\nu}$, so their transformation law involves twice as many Λ's:

$$F'_{\alpha\beta} = \Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu F^{\mu\nu}, \tag{264}$$

or, in matrix notation,

$$F' = \Lambda^T F \Lambda. \tag{265}$$

For the familiar boost matrix along the x axis, we get the humongous matrix equation

$$\begin{pmatrix} 0 & -E'_x & -E'_y & -E'_z \\ E'_x & 0 & -B'_z & B'_y \\ E'_y & B'_z & 0 & -B'_x \\ E'_z & -B'_y & B'_x & 0 \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & & \\ -\beta\gamma & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix} \begin{pmatrix} \gamma & -\beta\gamma & & \\ -\beta\gamma & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix}. \tag{266}$$

We'll leave the matrix algebra to scratch paper. The result is

$$E'_x = E_x \qquad B'_x = B_x \tag{267}$$

$$E'_y = \gamma(E_y - \beta B_z) \qquad B'_y = \gamma(B_y + \beta E_z) \tag{268}$$

$$E'_z = \gamma(E_z + \beta B_y) \qquad B'_z = \gamma(B_z - \beta E_y). \tag{269}$$

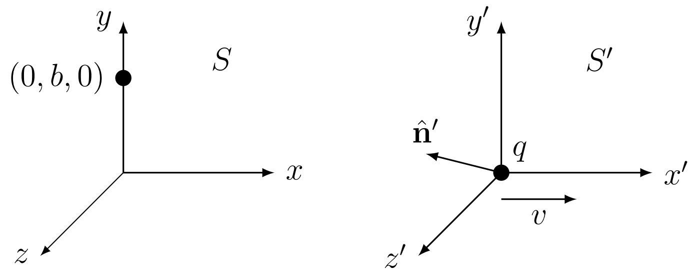
In total generality, an arbitrary boost transformation yields

$$\mathbf{E}' = \gamma(\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B}) - \frac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{E}) \tag{270}$$

$$\mathbf{B}' = \gamma(\mathbf{B} - \boldsymbol{\beta} \times \mathbf{E}) - \frac{\gamma^2}{\gamma + 1}\boldsymbol{\beta}(\boldsymbol{\beta} \cdot \mathbf{B}). \tag{271}$$

It is somewhat surprising that the components along the direction of the boost do not transform. But that's a rank two tensor for you.

To elucidate the physics behind this transformation, consider a frame S in which an observer sits at $(0, b, 0)$ watching a point charge moves along the x axis with velocity $\mathbf{v}$. Let S' be the rest frame of the charge and let $\mathbf{r}' = r'\hat{\mathbf{n}}'$ be the displacement in S' of the observer from the charge.



In S', the observer is at $(-vt', b, 0)$. The fields in S' at the observer is just a static Coulomb field

$$E'_x = \frac{-qvt'}{(r')^3} \qquad \text{and} \qquad E'_y = \frac{qb}{(r')^3}, \tag{272}$$

with $E_z = 0$ and $\mathbf{B} = 0$. To find the fields in S, we use the (inverse of the) transformations derived above (just switch the sign of β). We would also rather express the fields in S in terms of variables measured in S, i.e. not t' and r'. We can substitute these out using $(r')^2 = b^2 + v^2(t')^2$ and $t' = \gamma t$. We find

$$E_x = E'_x = -\frac{\gamma q v t}{\left[b^2 + (\gamma v t)^2\right]^{3/2}} \tag{273}$$

$$E_y = \gamma E'_y = \frac{\gamma q b}{\left[b^2 + (\gamma v t)^2\right]^{3/2}} \tag{274}$$

$$B_z = \gamma\beta E'_y = \beta E_y. \tag{275}$$

The electric field transverse to the motion has grown by a factor of γ and we've picked up a transverse magnetic field as well! Actually the magnetic field is not really so new, in the nonrelativistic limit $\gamma \simeq 1$, the field is

$$B_z = \frac{q}{c}\frac{vb}{r^3} \qquad \text{or} \qquad \mathbf{B} = \frac{q}{c}\frac{\mathbf{v} \times \mathbf{r}}{r^3}, \tag{276}$$

which is the Biot-Savart Law with $q\mathbf{v} = Id\boldsymbol{\ell}$.

Notice that $|B_z| = |E_y|$ as $v \to c$, and that the transverse component E_y dwarfs the longitudinal component E_x as $\gamma \to \infty$. The maximum transverse amplitude is $E_y = \gamma q/b^2$, while the maximum longitudinal amplitude is $E_x = O(q/b^2)$. Both components have a finite duration of order $\Delta t = O(b/\gamma v)$, so as $v \to c$, the observer sees a narrower pulse. Finally, note that the longitudinal pulse is odd while the transverse pulse is even, so if the observer time-averages over a period longer than $\Delta t \sim b/\gamma v$, then they won't see the longitudinal field at all.

Why just $F_{\mu\nu}F^{\mu\nu}$?

We now return to the question of why the Lagrangian of Maxwell electrodynamics is

$$\mathcal{L} = -\frac{1}{16\pi} F_{\mu\nu}F^{\mu\nu}, \tag{277}$$

without any additional terms. There's a purely classical answer, which is that its the only thing which gives us the principle of superposition! If we include in the Lagrangian additional terms that are higher than quadratic degree, then our equations of motion will not be linear in the fields. But perhaps you're willing to throw away the principle of superposition. Fortunately, there's a deeper answer. In fact, the real answer is that there usually are additional terms in the Lagrangian, but they're usually too small to see at the typical energies of classical electromagnetism, i.e. at energies around the eV scale.

To see this, work in units where $\hbar = c = 1$ and consider what units the coefficients in the Lagrangian must have. Recall that $[\hbar c]$ = [energy] $\cdot$ [length], so that [energy] = [length]$^{-1}$. Then the Lagrangian has units

$$[\mathcal{L}] = [\text{energy}]/[\text{length}]^{-3} = [\text{length}]^{-4}. \tag{278}$$

The electric field has units $[E]$ = [charge] $\cdot$ [length]$^{-2}$, but charge is unitless when $\hbar = c = 1$ (the fine structure constant in cgs is $e^2/\hbar c = 1/137$), so

$$[F_{\mu\nu}] = [E] = [\text{length}]^{-2}. \tag{279}$$

We can verify that the units work out in our Lagrangian (277):

$$\mathcal{L} = (\text{dimensionless }\#) F_{\mu\nu}F^{\mu\nu}. \tag{280}$$

What about more complicated terms? Gauge invariance and Lorentz invariance requires that our Lagrangian be a function of $F_{\mu\nu}F^{\mu\nu}$ or $F_{\mu\nu}\mathcal{F}^{\mu\nu}$. Let's focus on $F_{\mu\nu}F^{\mu\nu}$, and imagine $\mathcal{L}$ as a polynomial in $F_{\mu\nu}F^{\mu\nu}$. We could extend our Lagrangian to

$$\mathcal{L} = c_1 F_{\mu\nu}F^{\mu\nu} + c_2 (F_{\mu\nu}F^{\mu\nu})^2, \tag{281}$$

where $[c_2]$ = [length]^4. We could write $c_2 = 1/\Lambda^4$ where Λ is some fundamental energy scale—the energy scale of new physics beyond Maxwell's equations.

The simplest new physics we could think of might be quantum effects driving intermediate states. The lightest possible massive intermediate particle is the electron, so maybe $\Lambda \sim m_e = 0.5\,\text{MeV}$. In Fourier space,

$$F_{\mu\nu} = k_\mu A_\nu - k_\nu A_\mu, \tag{282}$$

so $F_{\mu\nu}F^{\mu\nu} = O(k^2)$. The new term in the Lagrangian therefore has a strength of order $(k/\text{MeV})^4$. But, for example, JILA works at $k \sim$ eV scales, which is too small to see the effects of the new physics. Modern terminology is to call $\mathcal{L} = c_1 F_{\mu\nu}F^{\mu\nu}$ an effective Lagrangian or an effective theory. It's not fundamental, but it has a certain range of validity where it looks fundamental.

6.4 Goldstone's Theorem and Spontaneous Symmetry Breaking

Mass of the Photon

"Mass of the Photon" is a poetic shorthand for a certain dispersion relation $\omega(k)$ of classical waves. If we have electromagnetism

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2\right)\mathbf{A} = 0 \tag{283}$$

and we assume $\mathbf{A} \sim e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$, then we find that $\omega = ck$. Putting in an $\hbar$ gives and energy-momentum dispersion relation $\hbar\omega = \hbar ck$, or $E = pc$. We've seen that $E = \sqrt{(pc)^2 + (mc^2)^2}$, so $E = pc$ is the dispersion relation of a massless particle.

Imagine that we instead have

$$\left(\frac{\partial^2}{\partial t^2} - \boldsymbol{\nabla}^2 + m^2\right)\Phi = 0, \tag{284}$$

where we've set $\hbar = c = 1$. For a plane wave $\Phi \sim e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)}$ this implies implies $\omega = \sqrt{k^2 + m^2}$, or with $\hbar$'s and c's:

$$(\hbar\omega)^2 = (\hbar ck)^2 + \left(mc^2\right)^2 \quad \Longrightarrow \quad E^2 = p^2 + m^2. \tag{285}$$

This equation of motion comes from a Lagrangian

$$\mathcal{L} = \frac{1}{2}\partial_\mu\Phi\partial^\mu\Phi - \frac{1}{2}m^2\Phi^2 - J\Phi \tag{286}$$

in the special case $J = 0$. Note that without the time derivatives, we have the static Yukawa solution

$$\Phi = \frac{e^{-mr}}{r}. \tag{287}$$

The Maxwell action is a complicated combination of $\partial_\mu A_\nu - \partial_\nu A_\mu$, but there is no A_μ^2 term. This gives us a massless photon. In fact, gauge invariance requires that $\mathcal{L}$ is a function of $F_{\mu\nu}$. A term like $A_\mu A^\mu$ is not gauge invariant, so we can say gauge invariance prevents a photon mass.

But we might want to think of massive photons. We might want to test that they're massless, or we might want to consider that massive "photons" actually do exist! They exist in the Meissner effect for superconductors and they exists as W and Z bosons. It takes a high price to throw away gauge invariance. Doing so in a naive way can also throw out charge conservation, which we might not want to do.

We'll obtain a theory with massless photons in a modern and not naive way. One element we'll need is Goldstone's theorem, which initially doesn't seem to give us a photon mass. But then we'll talk about the Higgs effect, which gives the gauge bosons a mass.

Consider a Lagrangian

$$\mathcal{L} = \frac{1}{2}\sum_{j=1}^{n}(\partial_\mu\varphi_j)(\partial^\mu\varphi_j) - V(\varphi), \tag{288}$$

where the potential is

$$V(\varphi) = \frac{\mu^2}{2}\varphi^2 + \frac{\lambda}{4!}\varphi^4, \tag{289}$$

and

$$\varphi^2 = \sum_{j=1}^{n}\varphi_j^2. \tag{290}$$

In the condensed matter literature, this is called a Ginzburg-Landau theory. In high energy theory, this Lagrangian models the Higgs field (without it's interactions).

What does the potential term do to the field equations? It appears as

$$-\frac{\partial\mathcal{L}}{\partial\varphi} = \frac{\partial V}{\partial\varphi}, \tag{291}$$

which is a force, just as in classical mechanics. It tries to drive φ to where V is minimized. Assuming V has a minimum, we linearize the equations of motion around the minimum, so $\mathcal{L}$ is approximately quadratic around the minimum. We write

$$V(\varphi) = V(\varphi_0) + \frac{1}{2}V''(\varphi_0)(\varphi-\varphi_0)^2 + \cdots \tag{292}$$

where we ignore all the nonquadratic pieces. Then the Lagrangian is (throwing away the constant piece)

$$\mathcal{L} = \frac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - V(\varphi) \tag{293}$$

$$= \frac{1}{2}\partial_\mu(\varphi-\varphi_0)\partial^\mu(\varphi-\varphi_0) - \frac{1}{2}V''(\varphi_0)(\varphi-\varphi_0)^2 \tag{294}$$

We define $\mu^2 = V''(\varphi_0)$ and identify it as a mass. This generalizes our notion of a mass to mean the second derivative of an arbitrary potential at its minimum. There's a corresponding dispersion relation $\omega^2 - k^2 - \mu^2 = 0$.

Let's consider a "hierarchy of models" where we play with this. First take $n = 1$ and a potential

$$V(\varphi) = \frac{1}{2}\mu_0^2\varphi^2 + \frac{\lambda}{4!}\varphi^4. \tag{295}$$

What is the mass and dispersion relation? We want the potential to go to infinity as $|\varphi| \to \infty$, so we require $\lambda > 0$. If $\mu_0^2 > 0$. We minimize V with $\varphi_0 = 0$, so the mass is $\mu^2 = \mu_0^2$

We can also consider $\mu_0^2 < 0$ (we just wrote it as a square for looks) . This is relevant because, for example, for a ferromagnet $\mu_0^2 \propto T - T_c$, so both cases for μ_0^2 are relevant. Then

$$0 = \frac{\partial V}{\partial\varphi} = \mu_0^2\varphi_0 + \frac{\lambda}{6}\varphi_0^3 = 0, \tag{296}$$

so the potential is minimized for $\varphi_0^2 = -6\mu_0^2/\lambda$. We see that there are two values φ_0 and $-\varphi_0$ which minimize the potential. The second derivative of the potential at $\varphi_0^2 = -6\mu_0^2/\lambda$ is

$$\frac{\partial^2 V}{\partial \varphi^2} = \mu_0^2 + \frac{\lambda}{2}\varphi_0^2 = -2\mu_0^2, \tag{297}$$

so the mass squared is $-2\mu_0^2$. There is a local maximum at $\varphi_0 = 0$ at which we have a negative second derivative of the potential, and hence an instability.

The original system had a discrete symmetry $\varphi(x,t) \to -\varphi(x,t)$. If the system minimizes its potential away from a symmetric point, we say "the symmetry is broken." If φ is in one potential minimum, then you can't tell that $\varphi \to -\varphi$ is a symmetry by looking at the excitations of the theory. To be more clear, let's write

$$\varphi(x,t) = \varphi_0 + \chi(x,t). \tag{298}$$

Then in terms of the excitations χ, the Lagrangian is

$$\mathcal{L} = \frac{1}{2}\partial_\mu \chi \partial^\mu \chi - V(\varphi_0 + \chi). \tag{299}$$

Writing out the potential, we find

$$V(\chi + \varphi_0) = \frac{1}{2}\mu_0^2(\chi + \varphi_0)^2 + \frac{\lambda}{4!}(\chi + \varphi_0)^4 \tag{300}$$

$$= \frac{1}{2}\mu_0^2\left(\chi^2 + 2\chi\varphi_0 + \cdots\right) + \frac{\lambda}{24}\left[\chi^4 + 4\chi\varphi_0^3 + 6\chi^2\varphi_0^2 + \cdots\right] \tag{301}$$

$$= \chi^2\left[\frac{1}{2}\mu_0^2 + \frac{\lambda}{4}\varphi_0^2\right] + 0 \cdot \chi, \tag{302}$$

where the linear term is zero after plugging in $\varphi_0^2 = -6\mu_0^2/\lambda$. If we wrote out the whole potential in terms of χ and φ_0, we would not be able to notice the symmetry $\chi + \varphi_0 \to -(\chi + \varphi_0)$, or $\chi \to -\chi - 2\varphi_0$.

6.5 3/18

Consider the field theory

$$\mathcal{L} = \sum \frac{1}{2}\partial_\mu \varphi_i \partial^\mu \varphi_i - V(\varphi).$$

Suppose $V(\varphi)$ has a minimum

$$\varphi(x,t) = \varphi_0 + \chi(x,t).$$

We acquire a mass

$$\frac{1}{2}m_\chi^2 = \left.\frac{\partial^2 V}{\partial \varphi^2}\right|_{\varphi_0}$$

Suppose $V(\varphi)$ had multiple degenerate minima. We pick one of the minimum energy configurations

$$\varphi(x,t) = \varphi_0 + \chi(x,t),$$

where φ_0 is one of the minima. If the original $V(\varphi)$ had a symmetry, it's hard to see in terms of χ. This is called spontaneous symmetry breaking. The "symmetry is broken" by the choice of φ_0. That's where we ended last time.

For our second example, consider $j = 2$, or φ is a complex field, so we have two real degrees of freedom. We'll consider a potential

$$V(\varphi) = \frac{1}{2}\mu_0^2\left(\varphi_1^2 + \varphi_2^2\right) + \frac{\lambda}{4!}\left(\varphi_1^2 + \varphi_2^2\right)^2.$$

The system has a continuous symmetry

$$\begin{pmatrix} \varphi_1' \\ \varphi_2' \end{pmatrix} = \begin{pmatrix} \cos\Omega & \sin\Omega \\ -\sin\Omega & \cos\Omega \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}.$$

A global rotation of φ leaves $\mathcal{L}$ unchanged. There is a conserved Noether current associated with this symmetry. If $\mu_0^2 > 0$, then V is concave up, the minimum of V is $\varphi_1 = \varphi_2 = 0$, so

$$\frac{1}{2}\frac{\partial^2 V}{\partial \phi_1^2}\bigg|_{\varphi_1=\varphi_2=0} = \frac{1}{2}\frac{\partial^2 V}{\partial \phi_2^2}\bigg|_{\phi_1=\varphi_2=0} = \frac{1}{2}\mu_0^2.$$

We can also check that

$$\frac{\partial\partial^2 V}{\partial\partial\varphi_1\partial\phi_2}\bigg|_{\varphi_1=\varphi_2=0} = 0.$$

If $\mu_0^2 < 0$ then we get a "Mexican hat." Defining $\varphi_1^2 + \varphi_2^2 = \rho^2$, the minimum of V is any point on the circle is $\rho^2 = -6\mu_0^2/\lambda$. We arbitrarily assume the symmetry breaks for

$$\varphi_1 = \varphi_0 = \sqrt{-\frac{6\mu_0^2}{\lambda}}, \varphi_2 = 0.$$

We define new fields that oscillate around the minima

$$\begin{aligned}
\varphi_1 &= \varphi_0 + \chi_1(x,t), \\
\varphi_2 &= \chi_2(x,t) \\
V(\varphi_1,\varphi_2) &= \frac{1}{2}\mu_0^2\left(\varphi_1^2 + \varphi_2^2\right) + \frac{\lambda}{24}\left(\varphi_1^2 + \varphi_2^2\right)^2 \\
&= \frac{1}{2}\mu_0^2\left[(\varphi_0 + \chi_1)^2 + \chi_2^2\right] + \frac{\lambda}{24}\left[(\varphi_0 + \chi_1)^2 + \chi_2^2\right]^2 \\
&= \frac{1}{2}\mu_0^2\left[\varphi_0^2 + 2\varphi_0\chi_1 + \chi_1^2 + \chi_2^2\right] + \frac{\lambda}{24}\left(\varphi_0^2 + 2\chi_1\varphi_0^2 + \chi_1^2 + \chi_2^2\right)^2
\end{aligned}$$

Working to quadratic order, we have

$$V = \chi_1\left[\varphi_0\mu_0^2 + \frac{\lambda}{6}\varphi_0^2\right] + \chi_1^2\left[\frac{1}{2}\mu_0^2 + \frac{\lambda}{24}\left(2\varphi_0^2 + 4\varphi_0^2\right)\right] + \chi_2^2\left[\frac{1}{2}\mu_0^2 + \frac{2\lambda}{24}\varphi_0^2\right].$$

Now we plug in $\varphi_0^2 = -6\mu_0^2/\lambda$. The linear term vanishes, and the rest is

$$V = \chi_1^2\left[\frac{1}{2}\mu_0^2 + \frac{\lambda}{4}\left(-\frac{6\mu_0^2}{\lambda}\right)\right] = -\frac{1}{2}\mu_0^2\chi_1^2.$$

The coefficient of the χ_2^2 term is also zero. Among the two separate excitations χ_1 and χ_2, only χ_1 is massive, while χ_2 is massless. This is an example of Goldstone's theorem, which states *When a global continuous symmetry is spontaneously broken, there is an accompanying massless mode.* When χ_2 is massless is equivalent to saying $\omega(k) \propto k$. The massless particle is called the ***Goldstone boson***. Most massless bosons are typically Goldstone bosons or gauge photons.

6.6 Higgs Effect

Electrodynamics appeared when we tried to promote a global symmetry to a local one. We had a symmetry

$$\phi(x) \to e^{i\Lambda(x)}\phi \quad \text{or} \quad \begin{pmatrix}\phi_1(x)\\ \phi_2(x)\end{pmatrix} = \begin{pmatrix}\cos\Omega(x) & \sin\Omega(x)\\ -\sin\Omega(x) & \cos\Omega(x)\end{pmatrix}\begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix}.$$

We needed to introduce the gauge field to soak up the extra terms we got from letting Ω be a function of x. (This is from Ryder QFT p. 301) We write the Lagrangian

$$\mathcal{L} = [(\partial_\mu + ieA_\mu)\phi][(\partial^\mu - ieA^\mu)\phi^*] + \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}.$$

We pick the sign of μ^2 so that $V(\varphi)$ is a sombrero. This is called the Abelian Higgs model. The minimum is at $|\phi| = \sqrt{\mu_0^2/2\lambda} \equiv a/\sqrt{2}$. Define

$$\phi(x) = \frac{a + \chi_1(x) + i\chi_2(x)}{\sqrt{2}}.$$

The potential is

$$\begin{aligned} V(\varphi) &= -\mu_0^2\left|\frac{a+\chi_1+i\chi_2}{\sqrt{2}}\right|^2 + \frac{\lambda}{4}\left(|a+\chi_1+i\chi_2|^2\right)^2 \\ &= -\frac{\mu_0^2}{2}\left(a^2 + 2a\chi_1 + \chi_1^2 + \chi_2^2\right) + \frac{\lambda}{4}(\cdot)^2 \\ &= \chi_1\cdot 0 + \chi_1^2\cdot(\text{nonzero}) + \chi_2^2\cdot 0, \end{aligned}$$

so just as before, we have a massive excitation and a massless Goldstone boson.

That was the potential, but now we look at the kinetic term.

$$\begin{aligned} D_\mu &= \partial_\mu - ieA_\mu \\ \varphi &= \frac{a + \chi_1 + i\chi_2}{\sqrt{2}} \\ D_\mu \varphi &= \frac{1}{\sqrt{2}}[\partial_\mu \chi_1 - ieA_\mu \chi_1 - ieA_\mu a + ie\partial_\mu \chi_2 - eA_\mu \chi_2] \\ &= \frac{1}{\sqrt{2}}[(\partial_\mu \chi_1 - eA_\mu \chi_2) + i(\partial_\mu \chi_2 - eA_\mu \chi_1 - eA_\mu a)]. \end{aligned}$$

We now square this to get $|D_\mu \varphi|^2$:

$$(D_\mu \varphi)^*(D_\mu \varphi) = \frac{1}{2}[\partial_\mu \chi_1 \partial^\mu \chi_1 + \partial_\mu \chi_2 \partial^\mu \chi_2] + \frac{1}{2}e^2 a^2 A_\mu A^\mu + \frac{2}{\sqrt{2}} eaA_\mu \partial^\mu \chi_2.$$

The second term here has a term for a massive photon. There's a connection between a broken symmetry and a massive vector boson.