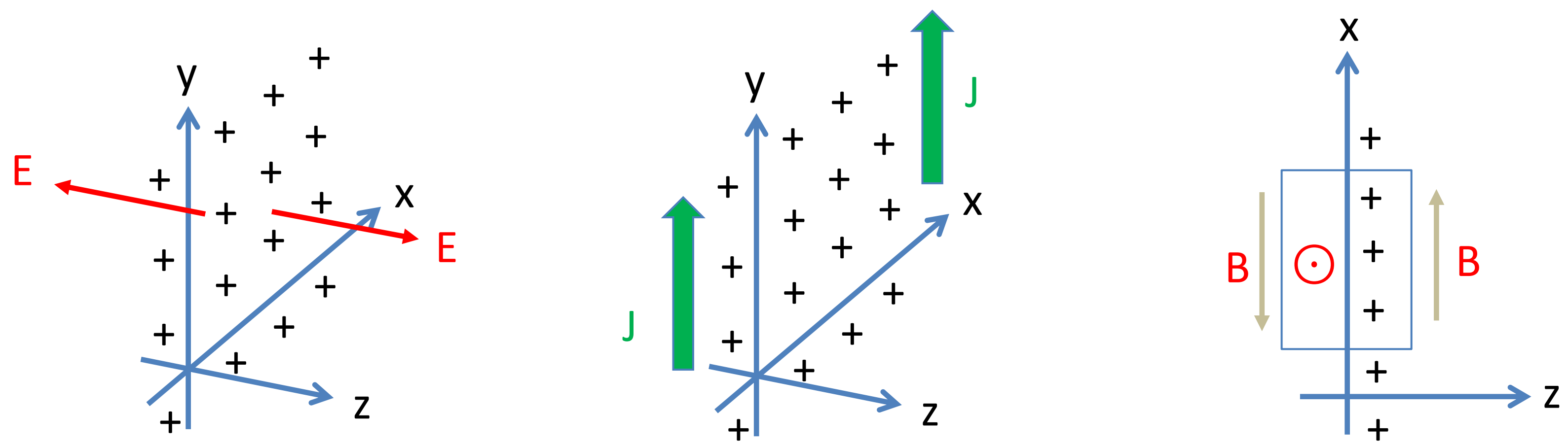
Electromagnetic radiation

Wave propagation is a subject that's heavily emphasized in undergraduate electricity and magnetism courses. But the waves come from somewhere and that is the topic of radiation. For physical insight let's begin with a discussion drawn from volume II of *The Feynman Lectures on Physics.* Imagine a uniform infinite (+) sheet of positive charge lying in the x-y plane. If it is stationary then Gauss's Law says there is a static electric field (red arrow) pointing along along the +z direction for z > 0 and along the – z direction when z < 0. There is no current so there is no *B* field so there are no waves produced by this charge distribution. A static field produces no radiation.

In the second figure the sheet is suddenly accelerated along the +y direction. There is now a sheet of *current* along +y shown by the green arrows. This generates a magnetic field *B*(t). To find *B* consider the rectangle whose sides are parallel to x and z and encloses a portion of the charge sheet. By Ampere's law,

$$\oint \vec{B} \cdot d\vec{l} = \mu_0 \oiint \vec{J} \cdot \overrightarrow{dA}$$

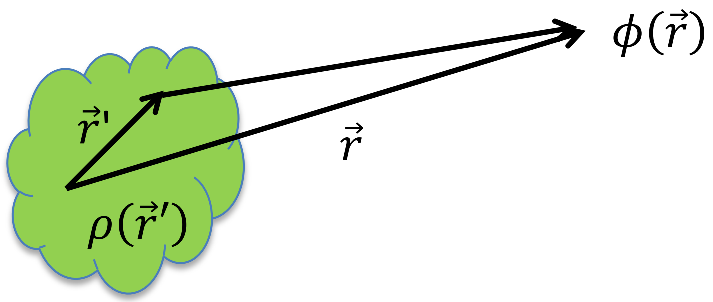
The line integral of B around the rectangle equals the flux of the current J through it. Since the sheets are infinite, B can only lie along the + or – x axis, as shown. *B* has changed from 0 to some non-zero value so it is a time-varying *B*-field so Faraday's law implies that a time-dependent *E* field is generated along the y-axis. That timy-varying *E* field now generates a *B* field, which in turn generates an *E* field and so on. The accelerated sheet of charge has generated *E* and *B* fields that obey the wave equation and propagate along the +/- z directions. The key ingredient for this classical radiation process is the *acceleration* of charge.



The dipole antenna

Radiation encompasses everything from microwave towers to gamma rays emitted from nuclei so we will need to focus on just a few examples. To begin, recall how to find the electric field produced by a time-independent distribution of charge, described by some charge density $\rho(\vec{r})$. The electric potential $\phi(\vec{r})$ is obtained by integrating over the charge distribution and then taking the gradient to obtain the electric field $\vec{E}$.

$$\phi(\vec{r}) = \frac{1}{4\pi\epsilon_0} \int \frac{\rho(\vec{r}')}{|(\vec{r} - \vec{r}')|}\, d\vec{r}' \qquad \vec{E} = -\nabla\phi \quad ,$$



Similarly, if there is some distribution of current described by a current density $\vec{J}(\vec{r})$ then we need to integrate over the region where this current exists to obtain the vector potential $\vec{A}$. The magnetic field $\vec{B}$ is then obtained by taking the curl of $\vec{A}$,

$$\vec{A} = \frac{\mu_0}{4\pi} \int \frac{\vec{J}(\vec{r}')}{|(\vec{r} - \vec{r}')|}\, d\vec{r}' \qquad \vec{B} = curl\ \vec{A}$$

Remember that $\vec{r}$ is the point of observation the potential and $\vec{r}$' points to the regions of charge or current density over which we need to integrate. You might also see the notation $d\vec{r}' = dx'dy'dz' = dV'$.

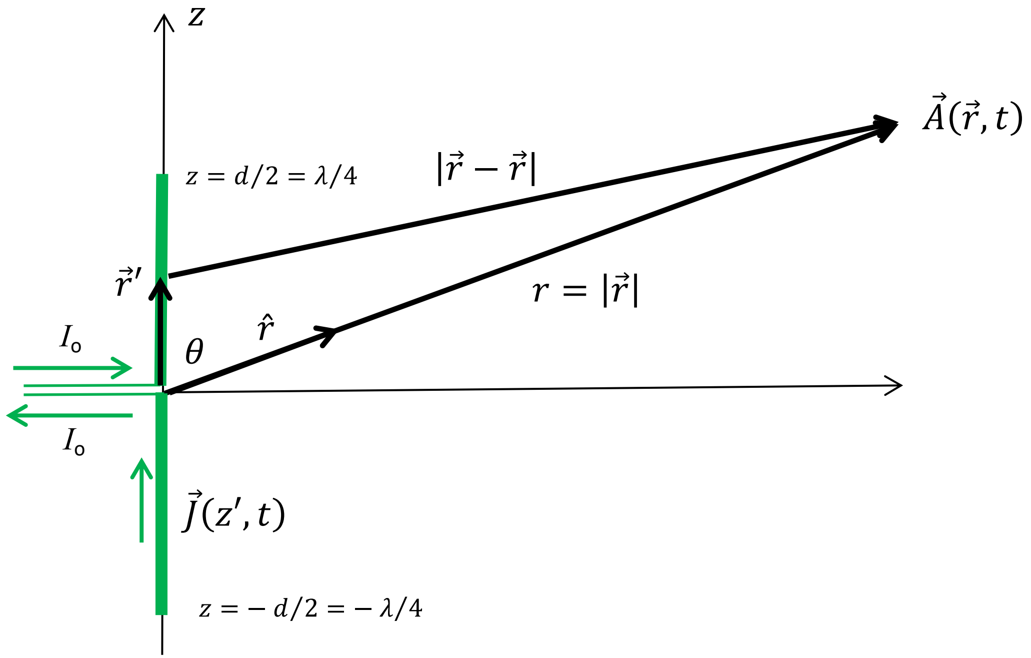
With static charge distributions it is often simpler to avoid the potentials and just use Gauss's Law or Ampere's law to obtain $\vec{E}$ or $\vec{B}$ directly. However, with radiation fields, it is mathematically *much* easier to first find the potentials and then take derivatives to get the electric and magnetic fields. And in quantum theory, the potentials are really the more fundamental quantities. Focus on the vector potential $\vec{A}$. The essential complication in finding the radiation fields is the finite speed of light *c*. The potential at point $\vec{r}$ and time *t* is determined by what the charges and currents were doing at $\vec{r}$' (as in the static case) but at an *earlier* time,

$$t_R = t - \frac{|(\vec{r} - \vec{r}')|}{c}$$

The quantity t_R is called the *retarded time*. It differs from the time *t* by the amount of time it takes light to go from point $\vec{r}$' (where the charge and current sources *are*) to the observation point $\vec{r}$. This makes the calculation much more difficult. You can look up the proof in any E&M textbook so I will just state the formal solution for the vector potential when the current density $\vec{J}(\vec{r},t)$ depends on time:

$$\overrightarrow{A}(\vec{r},t) = \frac{\mu_0}{4\pi}\int \frac{\vec{J}(\vec{r}',t_R)}{|(\vec{r}-\vec{r}')|}\,d\vec{r}' = \frac{\mu_0}{4\pi}\int \frac{\vec{J}\left(\vec{r}',t-\dfrac{|(\vec{r}-\vec{r}')|}{c}\right)}{|(\vec{r}-\vec{r}')|}\,d\vec{r}' \qquad \vec{B} = curl\,\vec{A}$$

This expression is similar the static case but now $\vec{J}(\vec{r}',t)$ is evaluated at the retarded time t_R , which *itself* depends on $\vec{r}$ and $\vec{r}$' . This integral is usually difficult to do but luckily there are many applications in which we care only about the fields that are very far away from the sources producing them and that greatly simplifies things.



We'll first look at the case of a *half-wave* antenna shown on the left. It's shown by thick green wires extending from z = 0 to d/2 and from z = 0 to -d/2 where $d = \lambda/2$. $\lambda = c/f$ is the wavelength of the radiation and $f = \omega/2\pi$ is the frequency. For FM radio at f = 100 MHz, $\lambda =$ 3 m so the antenna would be d = 1.5 m long.

The antenna is fed at z = 0 by a pair of wires carrying currents $I_o\,cos\omega t$ in and out respectively. The current on the antenna itself is maximum at z = 0 and goes to zero at each end of the antenna, so the current density along the antenna has the form,

$$\vec{J}\left(\overrightarrow{r'},t\right) = I_o\,cos\omega t\,cos(kz')\,\delta(x')\delta(y')\,\hat{z}$$

where $k = 2\pi/\lambda$. Charges slosh back and forth along the antenna at the driving frequency *f*. During half of the cycle the top end is positive and the bottom end is negative and then things reverse.

The observation point $\vec{r}$ is far away from the antenna so *r* >> *d*, λ . $\hat{r}$ is a unit vector parallel to $\vec{r}$ and $\vec{r}'$ points to places along the antenna where we need to integrate the current. From the figure you can see that $\vec{r}' = z'\hat{z}$. The fields close to the antenna (known as the *near zone*) are quite complicated but we will focus on the fields far away in what's termed the *far zone* or the *radiation zone*.

Since all the equations we use are linear, its permissible to represent sinusoidal quantities by complex exponentials and at the end of the calculation, take the real part of the answer to get the physical quantities. For example the current density can be written as $\vec{J}(\vec{r}', t) = I_o \, cos\omega t \, cos(kz') \, \delta(x')\delta(y') \, \hat{z} = I_o \, cos(kz') \, \delta(x')\delta(y') \, \hat{z} \, \mathrm{Re}\big(e^{-i\omega t}\big)$. The vector potential is now,

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi} \int \frac{\vec{J}(\vec{r}') \, e^{-i\omega\left(t - \frac{|(\vec{r}-\vec{r}\prime)|}{c}\right)}}{|(\vec{r} - \vec{r}')|} \, d\vec{r}' = \frac{\mu_0}{4\pi} \int \frac{I_o \, cos(kz') \, \delta(x')\delta(y') \, e^{-i\omega t} \, e^{i\omega \frac{|(\vec{r}-\vec{r}\prime)|}{c}}}{|(\vec{r} - \vec{r}')|} \, dx' \, dy' dz' \hat{z}$$

Next, we need to simplify $|\,\vec{r} - \vec{r}\,'|$. Out in the radiation zone $d \ll r$ so $\left|\vec{r'}\right| \ll |\vec{r}|$ so we can we can approximate $|\,\vec{r} - \vec{r}\,'|$:

$$|\,\vec{r} - \vec{r}\,'| = \sqrt{|\,\vec{r} - \vec{r}\,'|^2} = \sqrt{r^2 - 2\vec{r}\cdot\vec{r}' + r'^2} \approx \sqrt{r^2 - 2\vec{r}\cdot\vec{r}'} = \sqrt{r^2 - 2rz'\cos\theta} \approx r - z'\cos\theta$$

Using this approximation and substituting $\omega/c = k$,

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi} \int \frac{\vec{J}(\vec{r}') \, e^{-i\omega t} \, e^{i\omega \frac{|(\vec{r}-\vec{r}\prime)|}{c}}}{|(\vec{r} - \vec{r}')|} \, d\vec{r}' \approx \frac{\mu_0 I_0 \, e^{-i\omega\left(t - \frac{r}{c}\right)}}{4\pi} \int \frac{\cos kz' \; e^{-ikz'\cos\theta}}{r - z'\cos\theta} \, dz' \, \hat{z}$$

The physical field is obtained by taking the real part of this complex expression. Things can be further simplified because out in the radiation zone it's a good approximation to set $r - z'\cos\theta \approx r$ in the denominator. We are left with a one-dimensional integral over z' .

$$\vec{A}_{rad}(\vec{r}, t) \approx Re\left(\frac{\mu_0 I_0 \, e^{-i\omega\left(t - \frac{r}{c}\right)}}{4\pi r} \int_{-d/2}^{d/2} \cos kz' \; e^{-ikz'\cos\theta} \, dz' \right) \hat{z}$$

With some patience and trig identities this expression reduces to,

$$\vec{A}_{rad}(\vec{r}, t) = \frac{\mu_0 I_o \cos(kr - \omega t)}{2\pi k r} \, \frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin^2(\theta)} \, \hat{z}$$

It's a wave but with a peculiar angular dependence that is very important for signal communications.

Radiated Power

Given the vector potential we can now find $\vec{B}_{rad}$ and then $\vec{E}_{rad}$. The vector potential has been written things in terms of spherical coordinates *r* and θ. To find $\vec{B}_{rad} = curl\,\vec{A}$ the unit vector $\hat{z}$ needs to be in spherical coordinates: $\hat{z} = \hat{r}\cos\theta - \hat{\theta}\sin\theta$. Taking $curl\,\vec{A}$ there are terms proportional to 1/r, 1/r^2 and 1/r^3 . $\vec{B}_{rad}$ corresponds to the 1/r terms only. Why? The radiation fields are the ones that carry energy off to infinity. Remember that the Poynting vector $\vec{S}$ gives the energy per unit time flowing per unit area,

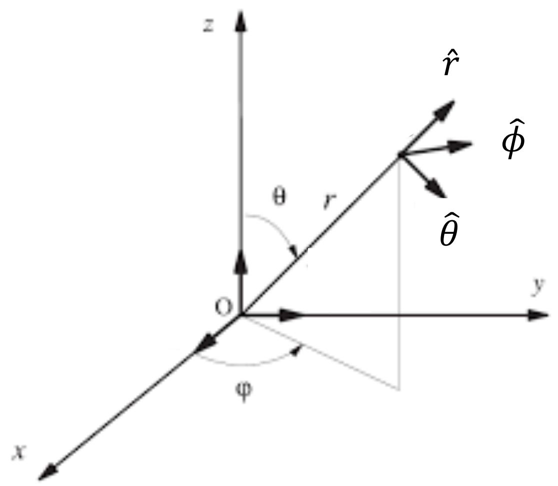
$$\vec{S} = \frac{1}{\mu_0}\,\vec{E}\;x\;\vec{B}$$

The total power radiated away is the integral of $\vec{S}$ over the surface of a sphere of radius *r* surrounding the antenna. Since the surface area of a sphere is $4\pi r^2$ then *S* must will vary as 1/r^2 in order for the power to be non-zero:

$$Power = \oint \vec{S}\cdot d\vec{A} \sim \frac{r^2}{r^2} > 0$$

This implies that the radiated fields $\vec{E}_{rad}$ and $\vec{B}_{rad}$ must both decrease as 1/r. Fields that decrease as 1/r^2 or 1/r^3 lead to a Poynting vector that decreases as 1/*r*^{4} or faster. The surface integral of $\vec{S}$ would then vanish as $r \to \infty$. Therefore no net power is radiated to infinity by such fields. They oscillate but they don't radiate. With that in mind, take $curl\,\vec{A}$ and keep only the part that varies as 1/r:

$$\vec{B}_{rad} = \frac{\mu_0 I_o \sin(kr-\omega t)}{2\pi r}\,\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin(\theta)}\,\hat{\phi}$$



The radiated magnetic field is azimuthal. In the far zone the electric field has magnitude $E_{rad} = cB_{rad}$ and is perpendicular to both the direction $\hat{r}$ of the radiated wave and to $\vec{B}_{rad}$:

$$\vec{E}_{rad} = c\vec{B}_{rad}\;x\;\hat{r} = c\;B_{rad}\;\hat{\theta}$$

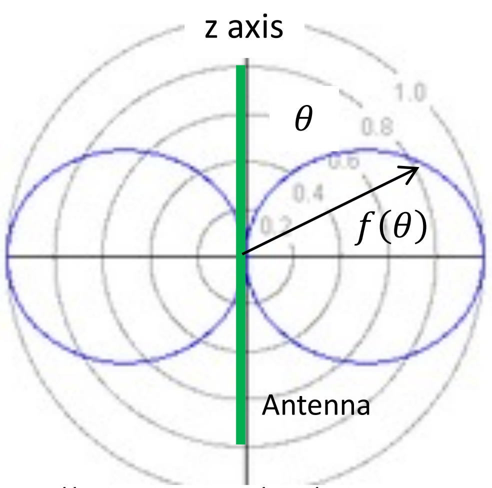
The unit vectors in spherical coordinates are shown in the figure.
The Poynting vector $\vec{S}$ points radially out along $\hat{r}$.

It's common to show the magnitude of the Poynting vector on a polar plot to emphasize the directionality of the antenna. Writing everything out we have,

$$\vec{S} = \frac{Z_0 I_0^2}{4\,\pi^2\, r^2}\, sin^2(kr - \omega t) \left(\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin(\theta)} \right)^2 \hat{r} \qquad Z_0 = \mu_0 c \approx 120\,\pi \approx 377\ \Omega$$

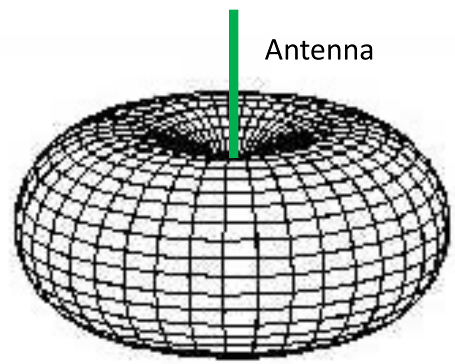
Taking the time average of $\vec{S}$ over one cycle converts the $sin^2(kr - \omega t)$ into a factor of ½. The time-averaged power per unit area radiated in the direction θ is now given by,

$$\langle \vec{S} \rangle = \frac{Z_0 I_0^2}{8\,\pi^2\, r^2} \left(\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin(\theta)} \right)^2 \hat{r} \ = \ \frac{Z_0 I_0^2}{8\,\pi^2\, r^2}\, f(\theta)\hat{r}$$



https://en.wikipedia.org/wiki/Dipole_antenna

The radiation pattern as a function of angle θ is shown on the polar plot. There is no power radiated along the axis of the antenna and maximum power radiated perpendicular to the antenna axis. The pattern doesn't depend on the azimuthal angle ϕ so it's symmetrical about the antenna axis. The radiated power pattern looks like a donut.



Since the dipole antenna radiation pattern is symmetrical around the z-axis, it's not very efficient. If you wish to send most of the energy in one direction then a more complicated antenna design is needed. Often, this involves an array of dipole antennas.

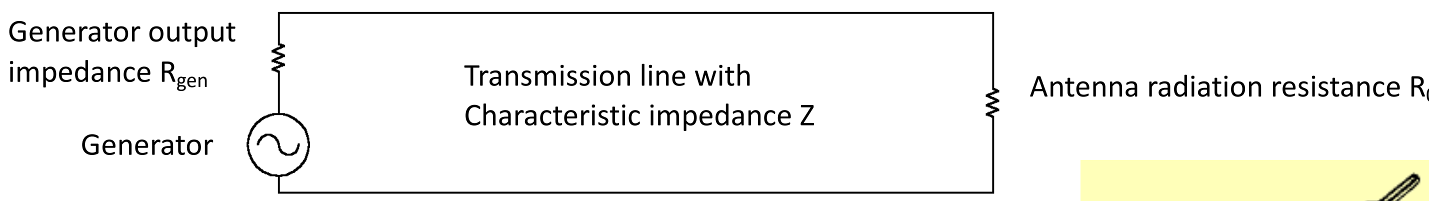
Radiation resistance

The time-average *total* radiated power is obtained by integrating $\langle\vec{S}\rangle$ over the surface of a sphere of radius *r*:

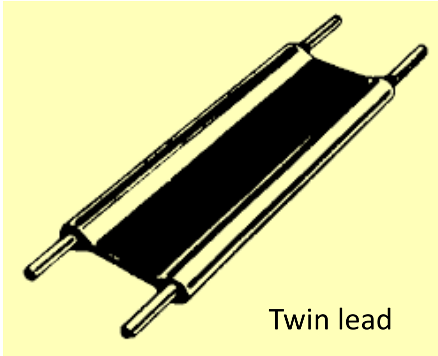
$$Average\ Total\ Power = \int \langle\vec{S}\rangle \cdot d\vec{A} = \int \frac{Z_0 I_0^2}{8\,\pi^2\, r^2} \left(\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin(\theta)}\right)^2 r^2 \sin\theta\, d\theta\, d\phi = \frac{Z_0 I_0^2}{8\,\pi^2}\, 2\pi \int_0^{\pi} \left(\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin(\theta)}\right)^2 \sin\theta\, d\theta\ = \frac{1}{2} I_0^2 R_0$$

The quantity R_0 is called the *radiation resistance* of the antenna. If you were to drive the same current $I_o\ cos\omega t$ through a resistor of value R_0 it would dissipate the same amount of power as the antenna radiates away. For the half-wave dipole antenna $R_0 \approx\ 73.09\ \Omega$. For other antenna shapes and sizes it will be different. This simple situation holds only at the resonant frequency of the antenna. If you drive it at a frequency which does not correspond to an integral number of half-wavelengths then the antenna acts like circuit involving R_0, capacitance and inductance.

Radiation resistance is important because we wish to radiate out the maximum power possible for a given generator voltage. The figure below shows the general idea. A voltage generator with an output impedance R_{gen} drives a transmission line with a characteristic impedance Z which is connected to an antenna that appears to the circuit as a resistance R_0. To radiate the maximum power for a given generator voltage, we need $R_{gen} = Z = R_0$. In other words, the generator is impedance-matched to the transmission line and the transmission line is impedance matched to the antenna.



You might think the coaxial cable would be the ideal transmission line but since its two conductors are not symmetrical, that can cause problems. An old favorite, going back to the early days of TV, is twin-lead. The spacing between the wires determines the characteristic impedance Z. A widely used value is Z = 300 Ohm. But that would be mismatched to a dipole antenna with $R_0 \approx\ 73.09\ \Omega$. This problem can be solved by the next scheme.

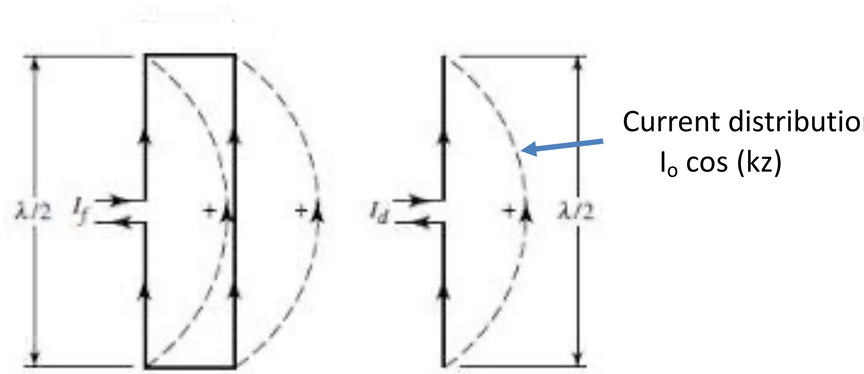


Folded dipole

The figure below shows a dipole antenna but with the two ends connected to each other by another section of wire. This is called a *folded dipole*. Remember that in the original dipole, the current has a cos(kz) shape (dotted line) which represents one half-wavelength. If we connect the wire between the ends, this represents a full wavelength round trip so the current in the folded part must be in the *same* direction as in the original dipole piece. That's like having two dipoles working in unison, so the fields will be twice as large for a given current. But that means 4 times as much radiated power for the same current. The radiation resistance is therefore 4 times as large,

$$R_{folded} = 4\,R_0 = 4\,x\,73.09\,\Omega = 292.4\,\Omega$$

This would be a better impedance match to the $Z = 300\,\Omega$ twin lead transmission line.

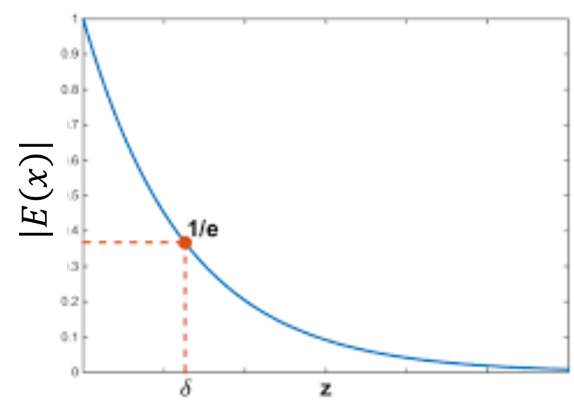


Folded dipole antenna   Ordinary half-wavelength dipole antenna

Ground Effect

Antennas here on the earth must contend with the ground beneath them. Depending on the soil composition and the amount of absorbed water, the ground is a moderately good conductor. Back in the 1800's the ground was actually used to carry the return current in telegraph systems. As you know, time-varying electric and magnetic fields only partially penetrate a conductor. by a characteristic length called the *skin depth* δ. This is the distance beneath the surface of a conductor surface where the electric and magnetic fields have fallen to 1/e of their values at the surface,

$$|E(x)| = |E(x=0)|e^{-x/\delta}$$

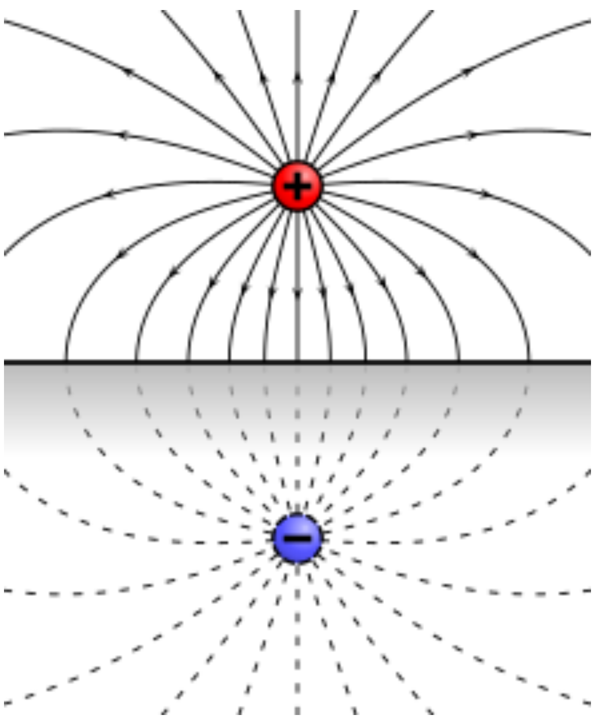


In good conductors (metals) the skin depth varies as,

$$\delta = \sqrt{\frac{\rho}{\pi \mu f}}$$

where ρ, μ, f are resistivity, permeability and frequency, respectively. At f = 100 MHz, the skin depth in copper $\delta = 6.5\ \mu m$. For poor conductors like the ground, the above formula holds for low frequencies but above 5 – 10 MHz the skin depth becomes independent of frequency and δ is a few meters. Nonetheless, it means that fields radiated by an FM radio antenna at 100 MHz are pretty much extinguished several meters below the ground. Suppose, as a very rough approximation, we treat the ground as a *perfect* conductor so the electric and magnetic fields of radio waves are *zero* beneath the surface. If we place an antenna above the ground, the E and B fields must satisfy the correct boundary conductions at the surface. That implies that right at the surface, the component of E *parallel* to the surface and the component of B *perpendicular* to the surface must be zero.

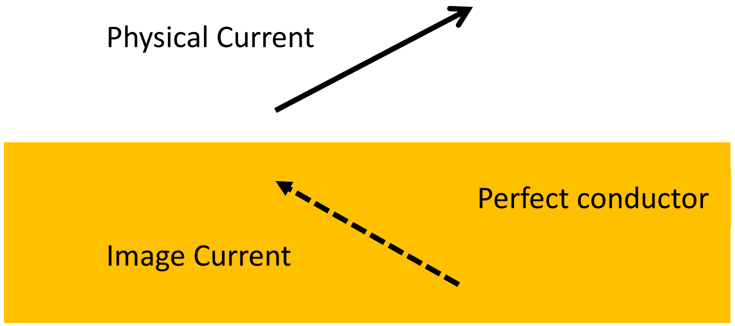
Images charges and currents

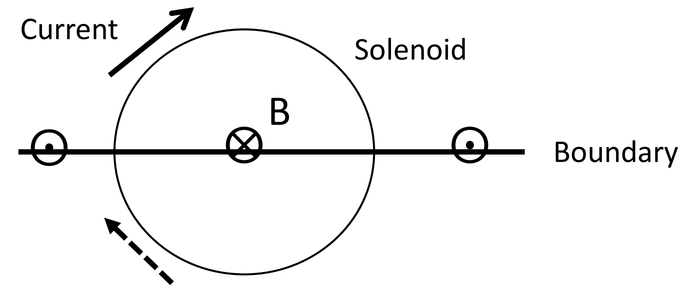


https://en.wikipedia.org/wiki/Method_of_image_charges

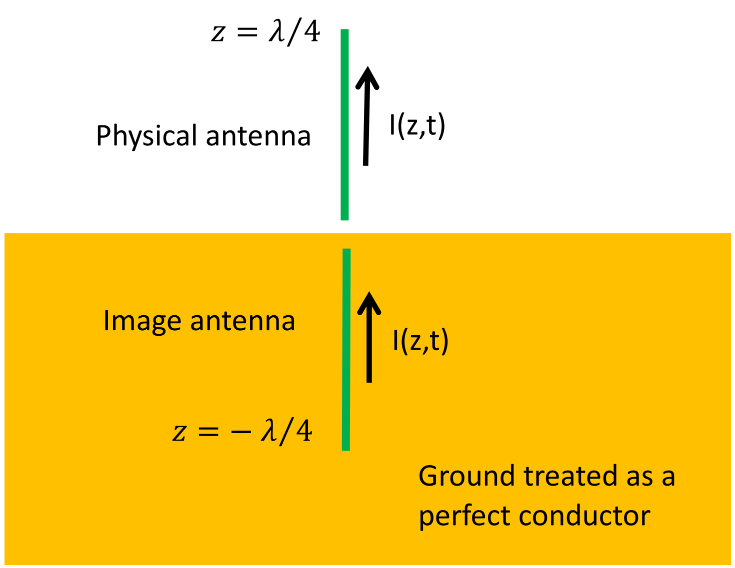
Problems like this can often be handled by the method of images, shown here for a point charge +q held above a perfect conductor. The conductor is replaced by an image charge –q located an equal distance below the boundary. The electric field from both charges together obeys the boundary conduction imposed by the conductor (parallel component of E = 0 at the surface). The physical electric field above the conductor surface is shown by the solid lines. It's just the field you would calculate from the two point charges.

A similar idea holds for currents. If we place an antenna carrying a current above the ground, the problem is to find an image *current* beneath the surface that makes the parallel component of E and the perpendicular component of B go to zero at the surface. The rule is that image currents parallel to the surface go in the opposite direction to those above the surface while image currents perpendicular to the surface go in the same direction as those above. The rule is shown in the next figure.





A quick way to understand this rule is to think of a long solenoid coming out of the page, carrying a clockwise current. The currents above and below the boundary have opposite parallel components and equal perpendicular components. This produces a magnetic field that is always parallel to the boundary and has *no* component perpendicular to the boundary. That's the same boundary condition that a perfect conductor would impose on the B-field. Now think of the lower half of the solenoid as the image current for the top half.

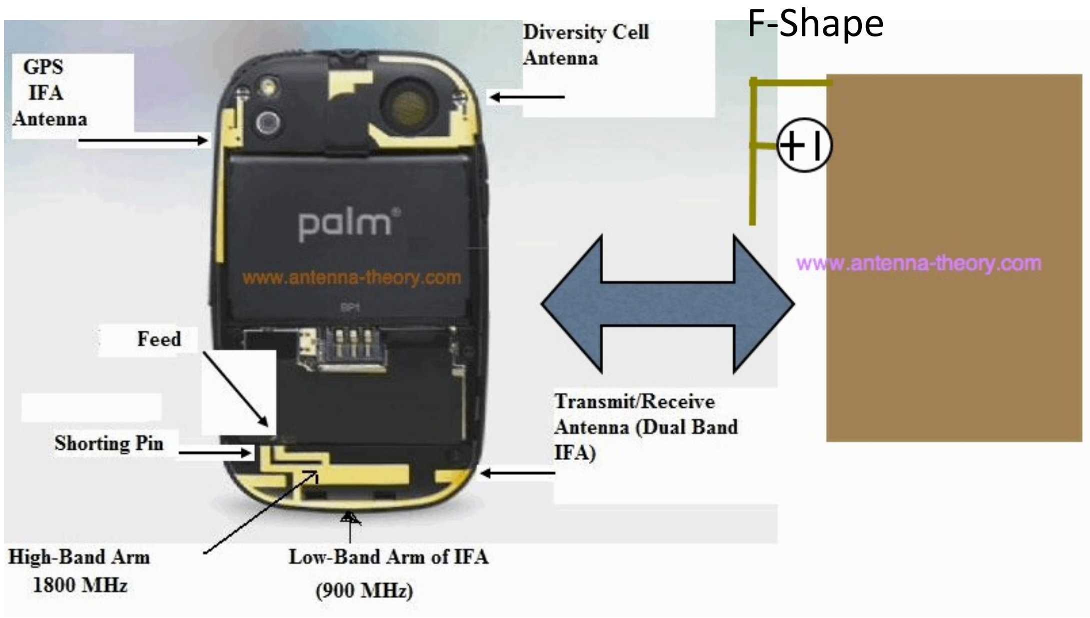


With that in mind, imagine we take our dipole antenna and cut it in half. The top half, still ¼ wavelength long, now sits above the ground, which is approximated as a perfect conductor. The effect of the ground on the E and B fields can be duplicated by an image of the top antenna, also ¼ wavelength long, carrying the *same* current in the vertical direction. For the same driving current, the radiation pattern should be identical to our original half-wave antenna stuck out in free space. Except of course that there are no E or B fields below the ground, so for the same antenna current, this antenna radiates only half the total power into space. It therefore has half the radiation resistance, namely R_{rad} = 36.5 Ohms.

The current generator would have one of its output wires connected to the antenna wire and the other half stuck in the ground. This arrangement is called a *grounded Marconi antenna*. Importantly, we have used the nearby conducting surface to reduce the size of the physical antenna from $\lambda/2$ to $\lambda/4$.

Cell Phones

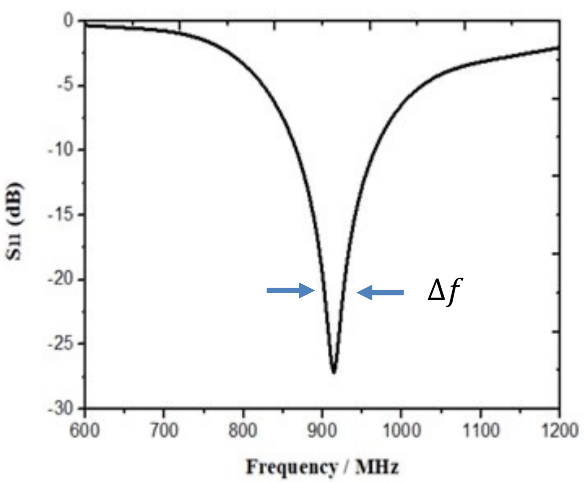
The size of hand-held phones is a major headache for antenna design. The simple half-wave dipole antenna sets the appropriate scale for an efficient antenna at a given frequency, namely $\lambda/2$. The Marconi antenna shows that a nearby conducting surface allows that to be cut in half, to $\lambda/4$. For example, the GPS antenna in the phone below operates at 1.575 GHz, corresponding to $\lambda/4 \approx 2$ ". Many such compact antennas are variations on the F-shape shown on the right. The large metallic rectangle is the ground plane – part of the phone structure itself. The circular (+ -) gap is the feed point where the driving currents enter and leave. The F shape is widely used since it allows some flexibility to vary the antenna length for the frequency band of interest. There are also antennas on this phone designed for 900 and 1800 MH. Interestingly GPS uses right-circularly polarized radiation but that kind of selectivity is difficult to design into a small space so this antenna settles for selecting out one linearly polarized component of the incoming GPS signal.



https://www.antenna-theory.com/design/gps.php

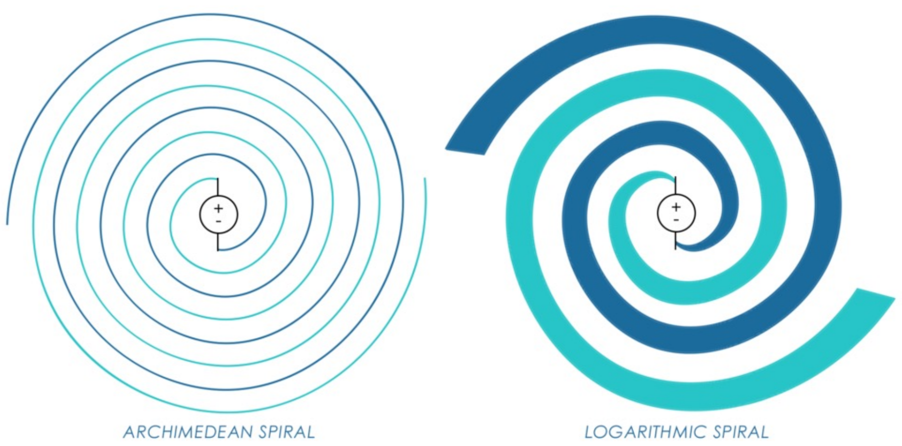
Bandwidth

The half-wave dipole antenna is a *resonant* system. Its peak efficiency occurs when driven at that frequency f for which the length of the antenna is $L = \lambda/2 = c/2f$ where c is the speed of light. At just this frequency the antenna appears like a pure resistance $R_0 = 73\ \Omega$ to the generator driving the current. If the generator and transmission line are matched to this value then no power is reflected back from the antenna toward the generator. It all goes into radiation. If, however, we transmit at a slightly different frequency then the antenna is no longer a pure resistance and some of the power is reflected back to the generator. For a $\lambda/2$ antenna resonant at about 900 MHz, the coefficient of reflection has the behavior shown below. It's a minimum right at resonance and increases as we move away from that frequency. The antenna becomes less efficient. The antenna *bandwidth* $\Delta f = f_2 - f_1$ is shown in the figure. f_1 and f_2 are often chosen to be the frequencies where the reflection coefficient has increased by $1/\sqrt{2}$ from its minimum value – basically a measure of the width of the resonance curve. Defined in this way, $\Delta f/f \approx 0.08$ for the $\lambda/2$ antenna.



https://www.researchgate.net/publication/344943034_Design_of_a_miniaturized_dipole_RFID_tag_antenna/figures?lo=1

A narrow bandwidth is desirable if you want to transmit or receive signals in a narrow range of frequencies and not interfere with anyone else. On the other hand, you might want an antenna that receives signals over a wide range of frequencies, often needed in military applications. The spiral antennas below are well-known broadband antennas which may have bandwidths of 10 GHz or more. They are also sensitive to circularly polarized microwaves.



https://jemengineering.com/blog-spiral-antennas/

Antenna Gain

Gain is a word we normally associate with amplifiers. For antennas, it concerns the angular dependence of the radiation. In the figure, the angular dependence that we worked out for the half-wave dipole antenna is shown with the two lobes described by the magnitude of the Poynting vector:

$$S_{dipole} = \frac{Z_0 I_0^2}{8\,\pi^2\,r^2}\left(\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{sin(\theta)}\right)^2 \qquad Total\ Power = \int S_{dipole}\ r^2 dr d\phi \sin\theta\, d\theta \ = \ \frac{1}{2} I_0^2 R_0$$

Now suppose the generator delivers the same amount of power $\frac{1}{2} I_0^2 R_0$ to a hypothetical *isotropic* antenna. (Hypothetical because there is no such antenna.) Since this hypothetical antenna sends the power out equally in all directions, *its* Poynting vector will be the total power radiated by the dipole divided by the area of a sphere:

$$S_{isotropic} = \frac{\frac{1}{2} I_0^2 R_0}{4\pi r^2}$$

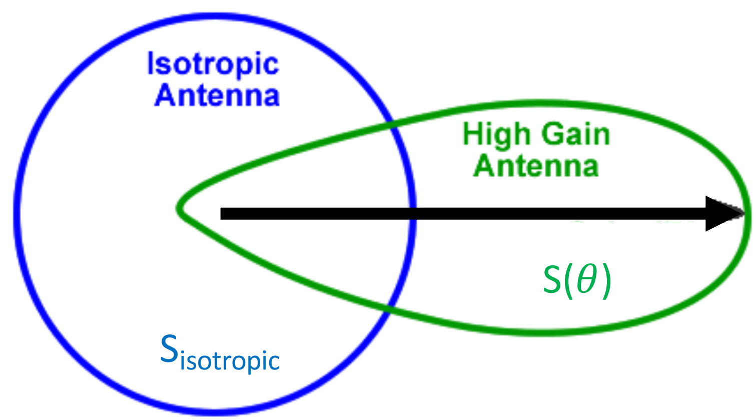
The *gain* of the real antenna is the ratio of the two Poynting vectors,

$$Gain(\theta) = \frac{S_{dipole}(\theta)}{S_{isotropic}} \qquad \frac{S_{dipole}(\theta = \pi/2)}{S_{isotropic}} = 1.64$$

Usually the antenna gain is taken to be its maximum value, which, for the dipole antenna, is 90 degrees relative to the antenna axis. In that direction the gain = 1.64. Engineers prefer the decibel unit, defined by the logarithm,

$$Gain\ (dB) = 10\ log_{10}\left(\frac{S_{dipole}(\theta)}{S_{isotropic}}\right)$$

The gain for an isotropic antenna is 0 dB. The maximum gain of the half-wave dipole antenna is 2.15 dB. For a highly directional antenna the gain can be much higher in the direction of maximum intensity, as shown.



https://www.ahsystems.com/articles/Understanding-antenna-gain-beamwidth-directivity.php

Antenna effective area

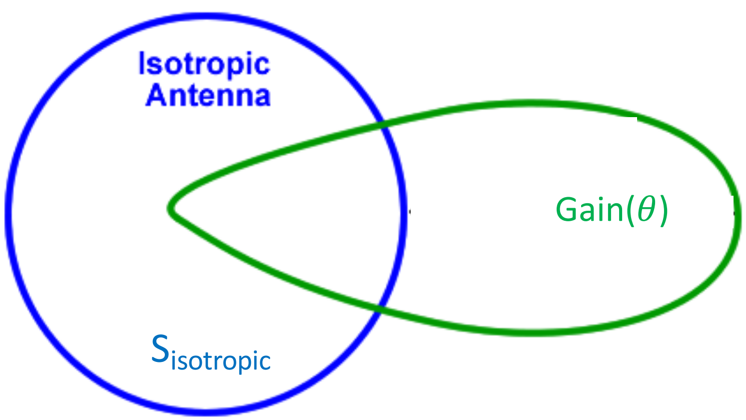
Antennas are of course used for both transmitting *and* receiving electromagnetic radiation. (Witness the bars on your cell phone.) But calculating the size of the signal received by an antenna is a *much* more difficult problem than calculating the power transmitted. Fortunately, nature let us off the hook. But we first need to define a new quantity call the *effective area* A_{eff} of an antenna. Suppose some radio station is beaming out radiation and there is distant antenna trying to receive it. Let *S* be magnitude of the transmitted Poynting vector at the receiving antenna and let $\langle P_{received} \rangle$ be the time-averaged power *extracted* from the incoming wave by the receiving antenna. Then the effective area of the receiving antenna is defined by,

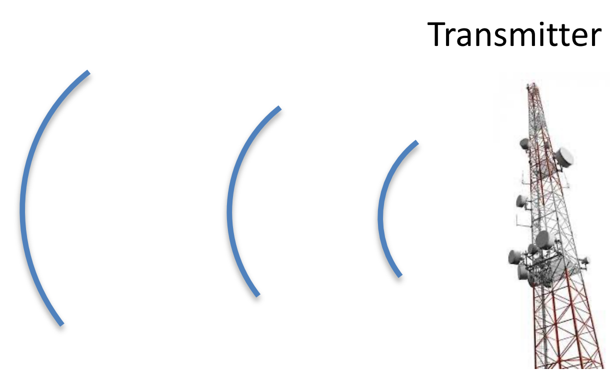
$$\langle P_{received} \rangle = \langle S \rangle \cdot A_{eff}$$

The surprising result is that for *any* antenna, so long as it doesn't dissipate power itself (from resistance in its conducting wires), the effective area and the gain are simply proportional, involving just the wavelength of the radiation being received:

$$\frac{A_{eff}}{Gain\ (maximum)} = \frac{\lambda^2}{4\pi}$$

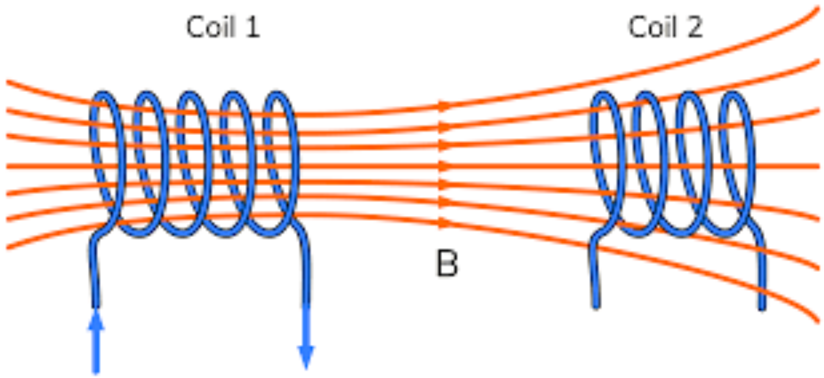
This equation is the connection between the transmitting and receiving properties of an antenna. We previously calculated the gain for a half-wave dipole antenna by working out the transmitted fields. But knowing the gain, the above equation gives us the effective area when we use the antenna as a receiver. For example, consider the antenna below. Its maximum gain is along the direction of the green lobe. But that's *also* the direction of its maximum effective area. As might seem intuitively obvious, to receive the strongest signal you should aim the green lobe at transmitter.





Reciprocity

The relationship between antenna gain and effective area is an example of what we call *reciprocity*. Subject to certain limitations on the transmitting medium, it's a general property of Maxwell's equations. For a more familiar example consider a transformer consisting of the two coils of wire.



Suppose we drive a current I_1 through coil 1 with a generator. This produces a time dependent magnetic field $B_1(t)$. Let Φ_{21} be the magnetic flux through coil 2 due to the field produced by coil 1. Then by Faraday's Law, the time rate of change Φ_{21} induces a voltage V_2 across its terminals,

$$V_2 = \frac{d\Phi_{21}}{dt} = M_{21}\frac{dI_1}{dt}$$

where M_{21} is the *mutual inductance*. It relates the flux generated in coil 2 *by* the current through coil 1.

Now turn things around and attach the generator to coil 2. This drives a current I_2 through coil 2 that produces a field $B_2(t)$ which in turn produces a time varying magnetic flux Φ_{12} through coil 1. By Faraday's law, this induces a voltage across the terminals of coil 1 given by,

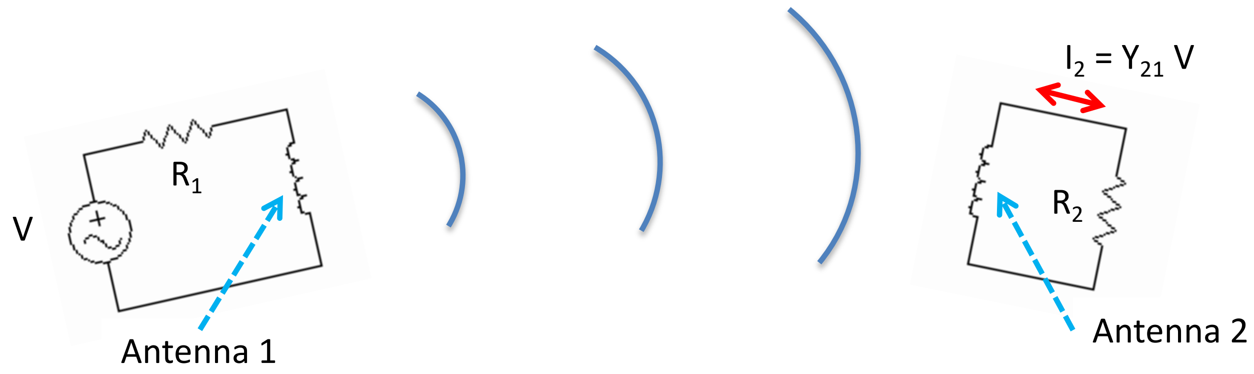
$$V_1 = \frac{d\Phi_{12}}{dt} = M_{12}\frac{dI_2}{dt}$$

where M_{12} relates the flux in coil 1 generated by a current in coil 2. Using the simple case of a solenoid wrapped around another solenoid, it's easy to show that $M_{12} = M_{21}$. But in fact this equality is true for *any* two coils and is an example of reciprocity. There's no radiation involved here since the transformers operate in the near zone. But the same idea holds true even if the mutual coupling were through the radiation fields, as in the next example.

Consider two scenarios.

Imagine there are two antennas radiation resistance R_1 and R_2 respectively. You can think of them as the two coils of a transformer that are very far apart, but in fact they could be any shape – dipole, coil, whatever. Each antenna is connected to an electrical circuit through which current can flow. To ensure impedance matching, each circuit includes a *real* resistor equal to its respective antenna resistance.

Case 1. Circuit 1 has a voltage generator V which drives a current through antenna 1 which transmits energy that is received by antenna 2. The energy received by antenna 2 generates a sinusoidally varying current of amplitude I_2 through circuit 2. Then $I_2 = Y_{21}$ V .



Case 2. Leaving everything else the same, now move the generator V so it drives a current through antenna 2. It becomes the transmitter and antenna 1 is the receiver. Antenna 1 picks up the incoming waves and generates a sinusoidally varying current of amplitude $I_1 = Y_{12}$ V through circuit 1.

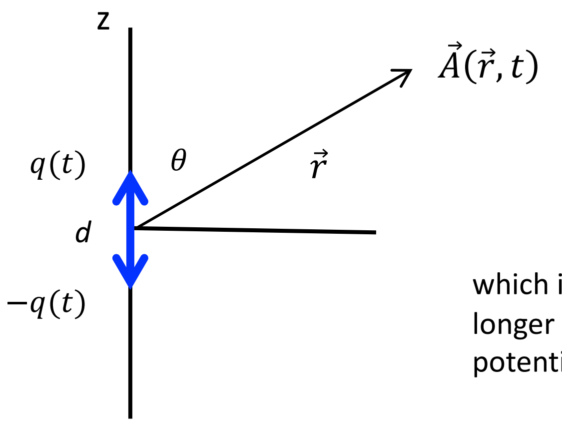
Reciprocity states that $Y_{12} = Y_{12}$. This statement is a generalization of the statement that for transformers $M_{12} = M_{21}$. It turns out there are similar reciprocity theorems for sound waves. A good reference can be found at https://www.cv.nrao.edu/~sransom/web/Ch3.html.

Hertzian dipole radiator

Antennas are ubiquitous in modern life but the electromagnetic radiation we mostly experience doesn't come from radio stations or cell phone towers but from atoms and molecules giving off visible light. But an atom is not a classical object and there is no obvious current generator driving the charge back and forth. In order to make a plausible connection to the quantum theory of radiation we'll first go back and rework the classical theory of radiation for a very small antenna whose size $d \ll \lambda$ where λ is the wavelength of emitted light. This is generally called a *Hertzian dipole* radiator. It contrasts with the hlaf-wave antenna for which $d = \lambda/2$. The inequality $d \ll \lambda$ if certainly well-satisfied for atoms, in which *d* < 1 nm and the shortest wavelength for visible light is 380 nm.

Imagine a charge $q(t) = Q \sin \omega t$ that sloshes back and forth from z = *d*/2 to z = −*d*/2. This amounts to a time-varying dipole moment given by $p(t) = d \cdot q(t) = Qd \sin \omega t = p_0 \sin \omega t$. The geometry is similar to the half-wave antenna but this dipole is so small that the current along z can be considered constant. The exact solution still has the same form as before:

$$\overrightarrow{A}(\vec{r},t) = \frac{\mu_0}{4\pi} \int \frac{\vec{J}(\vec{r}')\, e^{-i\omega\left(t - \frac{|(\vec{r}-\vec{r}')|}{c}\right)}}{|(\vec{r}-\vec{r}')|}\, d\vec{r}'$$



The charge sloshing back and forth corresponds to a current $I = dq/dt$ and a current density along the z-axis,

$$\vec{J}(\vec{r},t) = I(t)\, \delta(x)\delta(y)\, \hat{z} = \frac{dq}{dt}\, \delta(x)\delta(y)\, \hat{z} = Q\omega\, cos\omega t\, \delta(x)\delta(y)\, \hat{z}$$

which is similar to the expression for the current density of the half-wave antenna. However, we no longer have the $\cos kz'$ factor and we've replaced I_0 by $Q\omega$. Making those simplifications, the vector potential for the Hertzian dipole radiator is now,

$$\vec{A}_{Dipole}(\vec{r},t) \approx Re\left(\frac{\mu_0 \omega Q\, e^{-i\omega\left(t - \frac{r}{c}\right)}}{4\pi r} \int_{-d/2}^{d/2} e^{-ikz'\cos\theta}\, dz' \right)\hat{z}$$

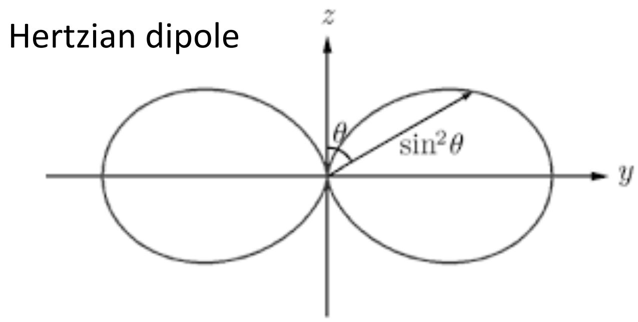
Since we're assuming $d \ll \lambda$ then $kz' \ll 1$ since z' in the integral never gets larger than *d*/2. Therefore we can replace the exponential inside the integral by 1. Writing the time derivative of the dipole moment as $\dot{p} = \omega dQ\, cos\, \omega t$ the vector potential is given by,

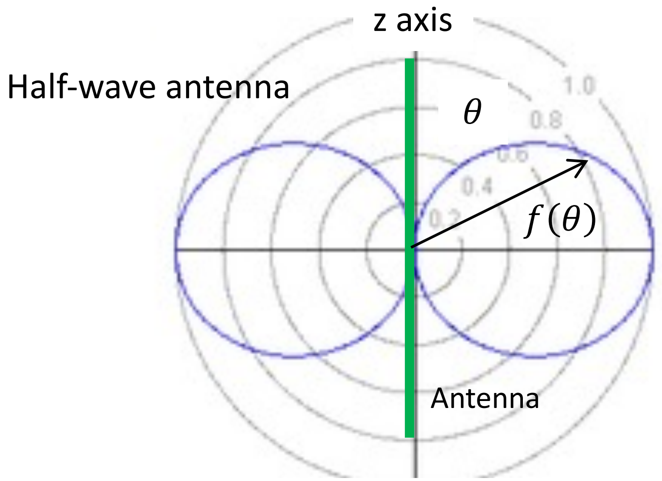
$$\vec{A}(\vec{r},t) = \frac{\mu_0}{4\pi r}\,\dot{p}\left(t - \frac{r}{c}\right)\hat{z}$$

Next, use $\vec{B} = curl\,\vec{A}$. Since *p* is evaluated at the retarded time, its argument includes *r*. Therefore, when taking the curl operation, the chain rule gives terms that depend on $\ddot{p}$, the *second* time derivative. As before, we care only about the radiation fields, which are the terms that at vary as 1/r. And in the radiation zone, $\vec{E}_{rad} = c\vec{B}_{rad}\,x\,\hat{r}$. Putting it all together we obtain the time-averaged Poynting vector,

$$\langle \vec{S} \rangle_{Hertz} = \left\langle \frac{1}{\mu_0}\,\vec{E}_{rad}\,x\,\vec{B}_{rad} \right\rangle = \frac{\mu_0}{16\,\pi^2\,c\,r^2}\,\langle\ddot{p}\rangle^2 sin^2(\theta)\,\hat{r} = \frac{\mu_0}{32\,\pi^2\,c\,r^2}\,\omega^4 p_0^2\,sin^2(\theta)\,\hat{r}$$

where $p_0 = Qd$. For the Hertzian dipole the angular dependence of the radiation is simply $sin^2(\theta)$. Although this is an entirely different function from the one derived for the half-wave antenna, the two radiation patterns actually look very similar. Both have a donut shape with a maximum in the $\theta = 90°$ direction. The maximum gain for the Hertzian dipole is 3/2.





The approximation we made to simplify the integral was,

$$e^{-ikz'\cos\theta} \approx 1 \qquad \lambda \ll max|z'|$$

This is generally called *dipole approximation*. Including higher order terms in the Taylor expansion is necessary for shorter wavelength radiation. We've introduced the Hertzian dipole radiator to motivate a brief discussion of radiation from atoms and molecules.

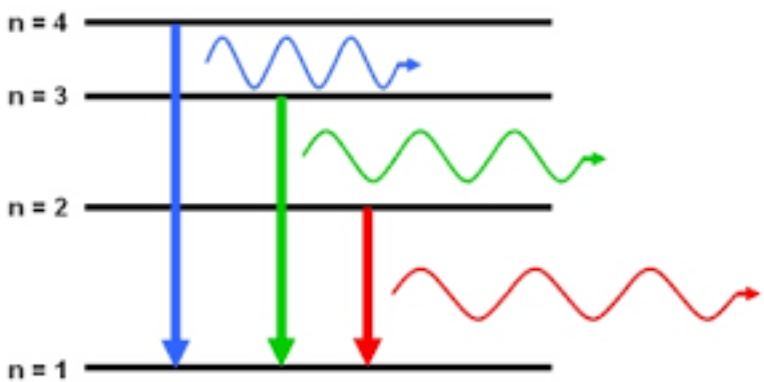
Radiation from quantum systems

Consider an atom sitting out in free space in a quantum state *n = 2* with energy E_2 . It can spontaneously jump to a state *n = 1* with energy E_1 and emit a *photon* whose frequency *f* obeys $hf = E_2 - E_1$ where $h = 6.626 \times 10^{-24}$ Joule-sec is Planck's constant. We've shown that the power radiated by a classical dipole is determined by the frequency and the dipole moment. The quantum mechanical equivalent is obtained by first thinking of the classical expression for the dipole moment of of a charge distribution,

$$\vec{p} = \int \vec{r}\,\rho(\vec{r}\,)\,d\vec{r}$$

where $\rho(\vec{r})$ is the charge density at point $\vec{r}$. You might think to replace $\vec{p}$ by its quantum mechanical expectation value where the charge density is replaced by $e\,|\psi(\vec{r})|^2$ which is the probability of finding an electron with wavefunction ψ at point $\vec{r}$ times its charge,

$$\langle\vec{p}\,\rangle = \int e\,|\psi(\vec{r})|^2\,\vec{r}\,d\vec{r}$$



https://astronomy.swin.edu.au/cosmos/e/emission+line

However this quantity is zero for atomic states like 1s, 2s, 2p, etc. since they have definite parity. But since we know that the radiation comes from transitions between atomic states, the relevant quantity for quantum radiation is the dipole moment operator *between* the initial and final quantum states,

$$\langle\vec{p}_{12}\,\rangle = \int e\,\psi^*_{final}\,\vec{r}\,\psi_{initial}d\vec{r}$$

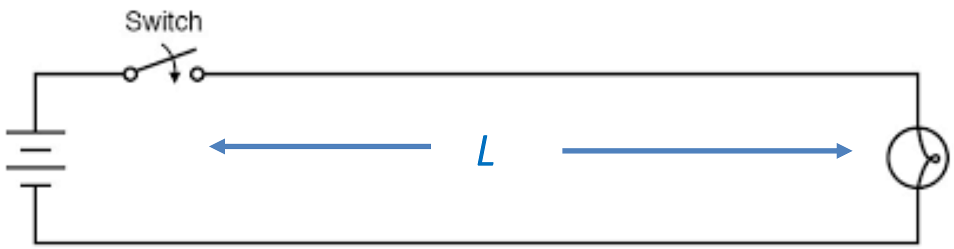
For example, the initial state might be a 2p and the final state would be 1*s*. The probability per unit time for an atom in an initial excited state to decay to a lower energy final state and emit a photon is given by,

$$\frac{Emissions}{sec} \propto \left|E_{initial} - E_{final}\right|^3 |\langle\vec{p}_{12}\,\rangle|^2 \propto \omega^3 |\langle\vec{p}_{12}\,\rangle|^2$$

This process is called *spontaneous emission* (to be distinguished from stimulated emission that occurs when we excite the atom with other photons.) The radiated power will be the rate of spontaneous emission times the energy of the emitted photon $\hbar\omega$ so, like the classical dipole radiator, the emitted power varies as $\omega^4|\langle\vec{p}_{12}\,\rangle|^2$. In classical radiation there is some "generator" that drives current back and forth on the antenna. In quantum mechanics the generator ultimately comes from quantum fluctuations of the electromagnetic field. Even with no applied fields ($\langle E\rangle, \langle B\rangle = 0$) quantum mechanics tells us that the ground state of the electromagnetic field has $\langle E^2\rangle, \langle B^2\rangle \neq 0$. The situation is analogous to a quantum harmonic oscillator in its ground state where position and momentum obey $\langle\, x\,\rangle, \langle\, p\,\rangle = 0$ but $\langle\, x^2\,\rangle, \langle\, p^2\,\rangle \neq 0$. The quantum theory of the electromagnetic field is a fascinating subject but beyond the scope of these lectures.

Transmission Lines

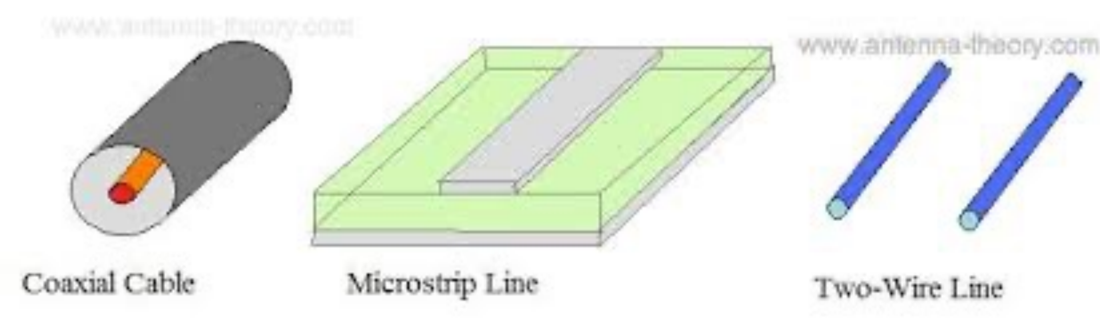
In electronics, we initially focus on *lumped* circuits, in which all the electromagnetic fields are confined (*lumped*) inside circuit elements such as resistors, capacitors, inductors, batteries and so on. The assumption is that the circuit components and the wires connecting them have negligible spatial extent. But obviously that can't always be true because the speed of light c = 3 x 10^8 m/sec is finite. In the circuit below, suppose L = 1 meter. When you close the switch it takes the light bulb $t = L/c \approx 3$ nsec to find out about it. If you were just turning on a light bulb you wouldn't notice a 3 nsec delay but if the switch and bulb were replaced by high-speed digital circuits then a 3 nsec delay would be easy to observe.



Or, look at it in the frequency domain. The assumption of lumped circuit analysis is that the size L of the circuit is much less than the free-space wavelength of light corresponding to the frequency at which the circuit operates,

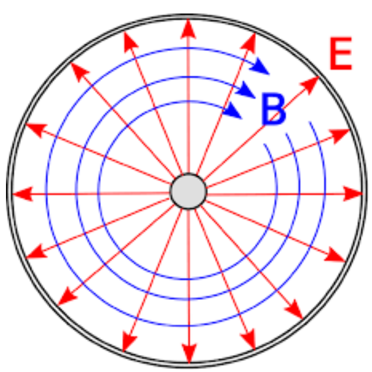
$$L \ll \lambda = \frac{c}{f}$$

For a DC circuit $f = 0$ so this inequality is satisfied, but if the battery were replaced by a sinewave generator running at f = 100 MHz (i.e., FM radio frequencies) then $\lambda = 3\ m$ and that is comparable to L. In that case, wave propagation along the connecting wires becomes important. In either case, we need to treat the wires connecting the generator to the load as a *distributed* circuit. That's the subject of transmission lines, the most common example being the ubiquitous coaxial cable.



We will focus on transmission lines like the ones shown on the left. All three consist of two metallic conductors separated by an insulator which could be vacuum or a dielectric. All three share one property – a constant cross-sectional shape in any plane perpendicular to the direction of propagation.

The coaxial cable is used in probably every physics lab in the world and is the easiest to analyze. Microstrip lines are used throughout printed circuitry. Two-wire lines were used back in the early days to connect the TV to the antenna for receiving network broadcasts. In each case an electromagnetic wave can propagate along the line. Its time-dependent electric and magnetic fields exist in the space localized around the conductors. The specific **E** and **B** field configurations will depend on the specific transmission line.



The figure shows the E and B fields in a cross-section of a coax cable. Imagine the coax inner conductor has radius a and uniform positive charge per unit length. The outer conductor has an inner radius b and carries an equal and opposite negative charge per unit length. Gauss's law leads to a radial electric field **E** and a voltage V between inner and outer conductor. Using Q = CV, It's easy to see that the coax will have a capacitance per unit length C_0 given by,

$$C_0(Farads/meter) = \frac{2\pi\epsilon_0}{ln\frac{b}{a}}$$

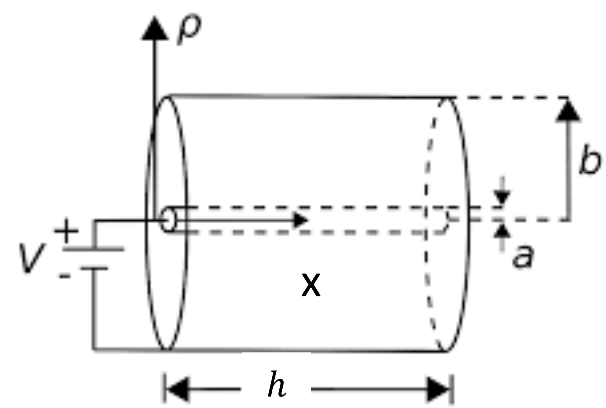
The coax will have an inductance per unit length L_0. To find that, imagine a current *I* coming out of the inner conductor and use Ampere's law to find **B**. To get the inductance per unit length use the energy stored in the **B** field for a length *h* of transmission line,

$$\frac{1}{2\mu_0}\int B^2\,dxdydz = \frac{1}{2}L_0 h\,I^2 \qquad \rightarrow \qquad L_0(Henries/meter) = \frac{\mu_0}{2\pi}\,ln\frac{b}{a}$$

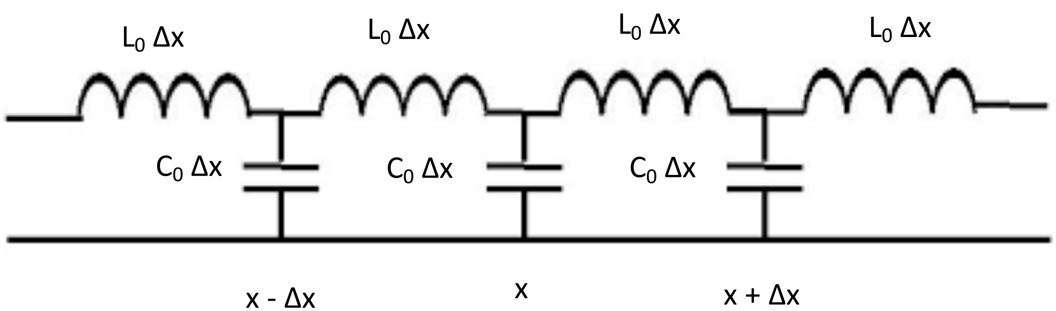
To find the inductance and capacitance per unit length for an arbitrarily-shaped transmission line you'll need to solve Maxwell's equations to get **E** and **B** and then relate them to the voltage and current via the energy equations:

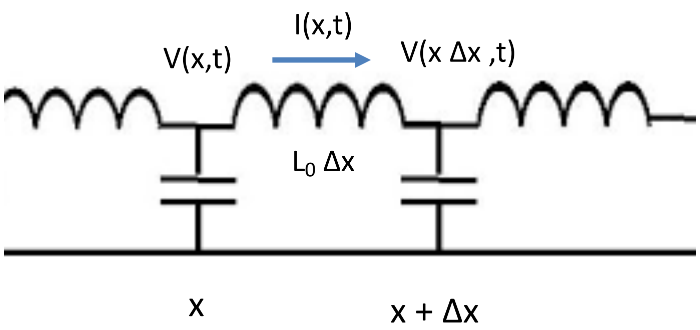
$$\frac{\varepsilon_o}{2}\int E^2\,dxdydz = \frac{1}{2}C_0 h\,V^2 \qquad \frac{1}{2\mu_0}\int B^2\,dxdydz = \frac{1}{2}L_0 h\,I^2$$

We will simply assume that for any transmission line there will be *some* C_0 and L_0 and go from there. As we will soon show, wavelike solutions for **E** and **B** will propagate back and forth in the x direction. Having given us C_0 and L_0, **E** and **B** have served their purpose and we'll deal only with the voltage V (x,t) between the two conductors and I (x,t) which is the current flowing through the inner conductor. Unlike the situation in lumped circuits, V and I now depend on *space* as well as time.



To understand the details of the propagation we'll treat the transmission line as a series of infinitely small capacitors and inductors. Each segment of the line with length Δx has a capacitance C_0 Δx and inductance L_0 Δx. For each of these infinitesimal *lumped* circuits we can apply Kirchoff's laws. This treatment can be found in many places, but I always recommend *The Feynman Lectures on Physics*, Vol. 2.





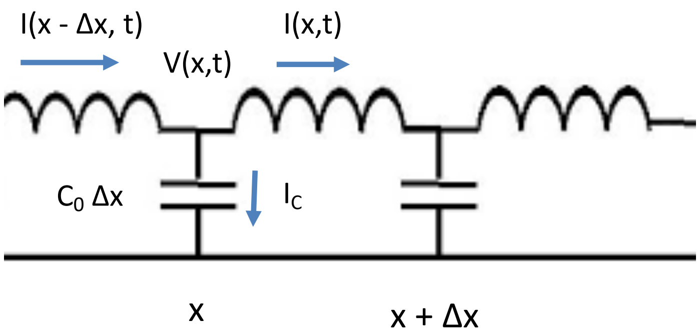
Focus on the little inductor between x and x + Δx. The voltage across it is given by,

$$V(x,t) - V(x+\Delta x,t) = L_0 \Delta x \, \frac{\partial I}{\partial t}$$

Expanding the left-hand side we have,

$$-\frac{\partial V}{\partial x}\Delta x = L_o \Delta x \, \frac{\partial I}{\partial t}$$

$$-\frac{\partial V}{\partial x} = L_o \, \frac{\partial I}{\partial t}$$



Next, focus on the capacitor $C_0 \Delta x$ located at x. Using current conservation, the current into the capacitor is,

$$I(x-\Delta x) - I(x) = I_C = C_0 \, \Delta x \frac{\partial V}{\partial t}$$

Expanding the left side gives,

$$-\frac{\partial I}{\partial x} = C_0 \frac{\partial V}{\partial t}$$

These two boxed equations are known as the the *Telegrapher's Equations*. They are essentially Faraday's Law and Ampere's Law in the context of transmission lines. Taking the space derivative of the first equation, the time derivative of the second and setting the mixed partial derivatives of *I* to be equal, we get the wave equation:

$$\frac{\partial^2 V}{\partial x^2} = L_0 C_0 \frac{\partial^2 V}{\partial t^2}$$

There is an identical wave equation for the current *I*(x,t). The transmission line supports wave-like solutions of *V* and *I* in which the phase velocity of the wave is given by,

$$\tilde{c} = \frac{1}{\sqrt{L_0 C_0}}$$

It can be shown that if the space surrounding the conductors of the transmission line is free of dielectrics, then $\tilde{c} = c$, the speed of light in a vacuum. Usually there is a dielectric around, in which case C_0 is proportional to the dielectric constant and the velocity is reduced accordingly. For coax cables, which typically have a Teflon-like dielectric between the inner and outer conductor, $\tilde{c} \approx 0.6\, c$.

Characteristic Impedance

Since V and I both obey the wave equation, which is linear, we can use phasor analysis to examine waves at a particular angular frequency. Represent the physical voltage and current along the line by the real part of phasors,

$$\mathrm{V}(x,t) = Re\left(\hat{V} e^{i(\omega t - kx)}\right) \qquad \mathrm{I}(x,t) = Re\left(\hat{I} e^{i(\omega t - kx)}\right)$$

To find *k*, substitute $e^{i(\omega t - kx)}$ into the wave equation, take the derivatives and cancel $e^{i(\omega t - kx)}$ from both sides:

$$\frac{\partial^2}{\partial x^2} e^{i(\omega t - kx)} = \frac{1}{\tilde{c}^2}\frac{\partial^2}{\partial t^2} e^{i(\omega t - kx)} \qquad \omega^2 = \tilde{c}^2 k^2 \quad \rightarrow \quad k = \pm\frac{\omega}{\tilde{c}}$$

For a given frequency ω (assumed to be positive) there are waves travelling to the right $(k = \omega/\tilde{c})$ and waves travelling to the left $(k = -\,\omega/\tilde{c})$. The complete solution on a transmission line generally involves 4 waves: right and left going voltage and right and left going current waves. But things simplify if we take a right-going voltage and current wave (denoted by a (+) subscript) and plug this solution into either one of the Telegrapher equations:

$$-\frac{\partial}{\partial x}\hat{V}_+ e^{i(\omega t - kx)} = L_o \frac{\partial}{\partial t}\,\hat{I}_+ e^{i(\omega t - kx)} \quad \rightarrow \quad k\,\hat{V}_+ = \omega L_0\,\hat{I}_+$$

Using the previous expression for the phase velocity, the ratio of the complex voltage amplitude to the complex current amplitude is given by,

$$\frac{\hat{V}_+}{\hat{I}_+} = \frac{\omega L_0}{k} = \frac{\omega L_0}{\omega/\tilde{c}} = L_0\tilde{c} = L_0\frac{1}{\sqrt{L_0 C_0}} = \sqrt{\frac{L_0}{C_0}} = Z_0$$

This ratio of complex amplitudes has the dimensions of Ohms and is called the *characteristic impedance*. If we now go through the calculation for a left-going wave we find that the characteristic impedance is $-Z_0$. Denoting the right and left going complex amplitudes with a (+) or (–) subscript we have,

$$\frac{\hat{V}_+}{\hat{I}_+} = Z_0 \qquad \frac{\hat{V}_-}{\hat{I}_-} = -Z_0$$

Knowing the complex voltage amplitudes for right and left going voltage waves, we automatically know the corresponding amplitudes of the current waves. The general solution at a given frequency is a sum of right and left-going waves,

$$V(x,t) = Re\left(\hat{V}_+ e^{i(\omega t - kx)} + \hat{V}_- e^{i(\omega t + kx)}\right) \qquad I(x,t) = Re\left(\hat{I}_+ e^{i(\omega t - kx)} + \hat{I}_- e^{i(\omega t + kx)}\right)$$

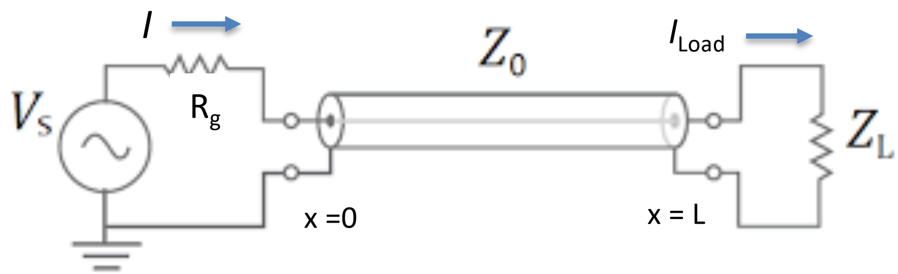
Using the characteristic impedance expressions, everything can be put in terms of just the right and left going voltage amplitudes,

$$V(x,t) = Re\left(\hat{V}_+ e^{i(\omega t - kx)} + \hat{V}_- e^{i(\omega t + kx)}\right) \qquad I(x,t) = Re\left(\frac{\hat{V}_+}{Z_0} e^{i(\omega t - kx)} - \frac{\hat{V}_-}{Z_0} e^{i(\omega t + kx)}\right)$$

To obtain the (+) and (-) voltage amplitudes will require two boundary conditions. Although Z_0 has the dimensions of Ohms, it is *not* a real resistance that dissipates energy! Our transmission line model contained no resistors (although a more realistic model would include them). Z_0 depends on the capacitance and inductance per unit length so it depends on the specific cross-sectional shape of the transmission line. Coax cables typically have $Z_0 \approx 50\ \Omega$. Two-wire lines such as old-fashioned antenna cable might have $Z_0 \approx 300\ \Omega$. Characteristic impedances typically vary from about 10 – 500 Ohms, depending on the transmission line geometry. The variation is not large because Z_0 usually varies logarithmically with the cable dimensions. For example, in coax cables Z_0 varies as $\ln(R_2/R_1)$ where R_1 and R_2 correspond to the radii of the inner and outer conductor, respectively. For our transmission line model Z_0 is *real* (i.e., resistive). If we were to include some actual resistance, along with L_0 and C_0, in the transmission line model, then Z_0 would acquire an imaginary part. However, for a great deal of high frequency electronics, treating Z_0 as real is accurate enough.

Boundary values

The next problem is to find $\hat{V}_+$ and $\hat{V}_-$. Focus on the circuit shown below. Without the transmission line this is just a voltage divider. The transmission line is a distributed circuit that connects the lumped circuits at x = 0 and x = *L*. The assumption in such schematics is that the lumped circuits on either end have no spatial extent and can be treated with Kirchoff's voltage and current laws. The transmission line piece has the general solution we just wrote down, with two unknown amplitudes. The circuits on either end provide the two boundary conditions.



Focus first on the situation at x = 0. The current from the generator must equal the total transmission line current at x = 0. For the transmission line current, set x = 0 in each exponential. As usual, all the $e^{i\omega t}$ factors cancel out and we have,

$$\hat{I} = \hat{I}_+ + \hat{I}_- = \frac{\hat{V}_+}{Z_0} - \frac{\hat{V}_-}{Z_0}$$

Using Kirchoff's voltage law at x = 0 we also have $\hat{V}_S - \hat{I}R_g = \hat{V}_+ + \hat{V}_-$. Before solving set $R_g = Z_0$. In other words, the generator output impedance is purposely made equal to the characteristic impedance of the transmission line, typically 50 Ohms. That is typical of most high frequency electronics. With that,

$$\hat{V}_+ = \frac{\hat{V}_S}{2}$$

This result has a simple physical interpretation. If $Z_0 = R_g$, then the transmission line divides the generator voltage by ½. If, for example, $R_g = Z_0$ = 50 Ohms, then the transmission line looks like a 50 Ohm resistor to the generator. However, this holds true *only* so long as there is only a right-going wave present at x = 0. That can happen if the generator produces a voltage step at t = 0. The step voltage travels down the line, reflects off the far end and travels back in a total time $T = 2L/\tilde{c}$. Until that reflected wave reaches x = 0, the generator doesn't know about it and the transmission line looks like a resistor of value Z_0. But after that, we need to add the reflected wave. We'll leave that as a homework problem.

We still need to solve for the left-going voltage amplitude. To do that, go to x = L. Again, the total current on the line at x = L must equal the current through the load impedance. Using Kirchoff's current law,

$$\hat{I}_+ e^{-ikL} + \hat{I}_- e^{ikL} = \hat{I}_{Load}$$

The total voltage on the line at x = L must equal the voltage across the load:

$$\hat{V}_+ e^{-ikL} + \hat{V}_- e^{ikL} = Z_L \hat{I}_{Load}$$

Solving for the ratio of the complex voltage amplitudes we find,

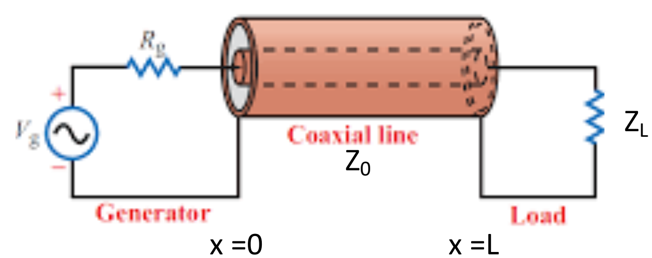
$$\frac{\hat{V}_-}{\hat{V}_+} = e^{-2ikL}\frac{Z_L - Z_0}{Z_L + Z_0} = e^{-2ikL}\,\Gamma_V$$

Γ_V is known as the voltage reflection coefficient. In general, Γ_V is complex but there is one important case where it is zero. That occurs when $Z_L = Z_0$, the characteristic impedance of the line in which case there is *no reflected wave*. This is called *terminating* the transmission line in its characteristic impedance. It's done everywhere in high frequency circuits to avoid reflections which cause standing waves and create problems. The transmission lines often have $Z_0 = 50\ \Omega$ so both the input and output impedance of amplifiers (and other components) are purposely made to be $50\ \Omega$. If such an amplifier is connected to the line at x = L there will be no reflected wave.

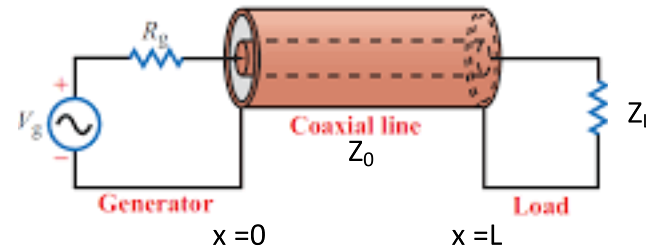
We now have everything to determine the currents and voltages in the circuit. The full solution is given by,

$$V(x,t) = Re\left(\hat{V}_+ e^{i(\omega t - kx)} + \hat{V}_- e^{i(\omega t + kx)}\right)$$

$$I(x,t) = Re\left(\frac{\hat{V}_+}{Z_0} e^{i(\omega t - kx)} - \frac{\hat{V}_-}{Z_0} e^{i(\omega t + kx)}\right)$$



Assume the generator is built with $R_g = Z_0$. Also use the expression for the ratio of the left to right going voltage waves. The voltage is given by,



$$V(x,t) = Re\left(\hat{V}_+ e^{i(\omega t - kx)} + \hat{V}_- e^{i(\omega t + kx)}\right) = Re\left(\frac{\hat{V}_S}{2}\, e^{i(\omega t - kx)} + \frac{\hat{V}_S}{2}\, \Gamma_V\, e^{-2ikL} e^{i(\omega t + kx)}\right)$$

$$V(x,t) = Re\left(\frac{\hat{V}_S}{2}\, e^{i\omega t}\left[e^{-ikx} + \Gamma_V e^{-2ikL} e^{ikx}\right]\right)$$

Similarly, the current is given by,

$$I(x,t) = Re\left(\frac{\hat{V}_+}{Z_0} e^{i(\omega t - kx)} - \frac{\hat{V}_-}{Z_0} e^{i(\omega t + kx)}\right) = Re\left(\frac{\hat{V}_S}{2\, Z_0}\, e^{i\omega t}\left[e^{-ikx} - \Gamma_V e^{-2ikL} e^{ikx}\right]\right)$$

We usually aren't probing along the transmission line to measure V and I at each x but instead, we care about the behavior at x = 0 and x = L where things are connected. For example, find the impedance looking into the line at x = 0. That's the ratio of the full complex voltage amplitude divided by the full complex current amplitude at x = 0:

$$Z(x=0) = \frac{\hat{V}(x=0)}{\hat{I}(x=0)} = \frac{\frac{\hat{V}_S}{2}\left[e^{-ikx} + \Gamma_V e^{-2ikL} e^{ikx}\right]}{\frac{\hat{V}_S}{2\, Z_0}\left[e^{-ikx} - \Gamma_V e^{-2ikL} e^{ikx}\right]} = Z_0 \frac{1 + \Gamma_V e^{-2ikL}}{1 - \Gamma_V e^{-2ikL}} = Z_0 \frac{Z_L + i\, Z_o \tan kL}{Z_0 + i\, Z_L \tan kL}$$

This last formula is particularly useful and we will consider three different cases.

1. Terminated Line: First consider the case $Z_L = Z_0$ in which case $\Gamma_V = 0$ and there will be no reflected wave. For any length of line, we have,

$$Z(x = 0) = Z_0 \qquad\qquad Z_L = Z_0$$

As stated earlier, when there is no reflected wave, then no matter what the length of the transmission it appears to the generator like an impedance Z_0 .

2. Half-wave line: Suppose the line is one half-wavelength long, $L = \lambda/2$. Then $k = 2\pi/\lambda$ so $k = 2\pi/2L$ and,

$$\tan kL = \tan \pi = 0 \quad \rightarrow \quad Z(x = 0) = Z_L$$

In other words, if the line has length $L = \lambda/2$ (or any integral multiple of that) then the generator thinks the load impedance is connected right at x = 0. Of course in the *time* domain, the signal still takes a finite time to reach the load. But in the steady state situation with a signal generator operating at constant frequency, the half-wave line is invisible.

3. Quarter-wave line: Now consider a line with length $L = \lambda/4$. Then $k = \pi/2L$ and we have,

$$\tan kL = \tan \frac{\pi}{2} = \infty \quad \rightarrow \quad Z(x = 0) = \frac{Z_0^2}{Z_L}$$

The transmission line inverts the load impedance. If a quarter-wave line is short-circuited at x = L then $Z_L = 0$ and,

$$Z(x = 0) = \infty$$

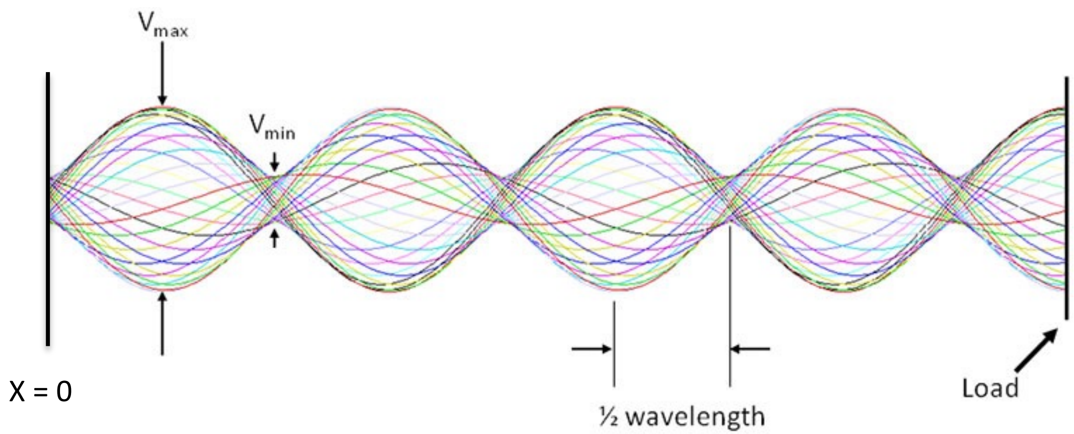
To the generator, the line looks like an *open* circuit! Similarly, if a quarter wave line is open-circuited at x = L then the generator thinks the line is short circuited. The quarter-wave line acts like an *impedance transformer*. If the load is capacitive then,

$$Z_L = \frac{1}{i\omega C} \rightarrow Z(x = 0) = Z_0^2 i\omega C = i\omega L \quad L = Z_0^2 C$$

The line has effectively transformed a capacitor into an inductor. The reverse would also be true.

Standing waves

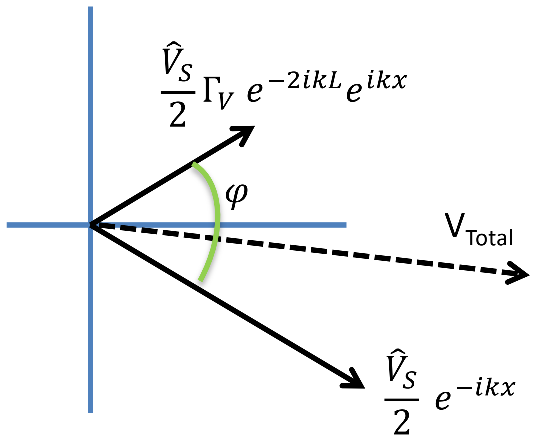
If right-going and left-going waves are both present on the line there will be standing waves. Assuming we have a sinusoidal generator voltage, the voltage on the line V(x,t) would look something like this figure where different colors indicate different times. At every point the voltage oscillates at ω and the amplitude of the oscillation varies sinusoidally with x. It's a standing wave.



From: nutsvolts.com

To derive the envelope function of the standing wave, use the solution for the voltage on the line:

$$V(x,t) = Re\left(\frac{\hat{V}_S}{2}\, e^{i\omega t}\left[e^{-ikx} + \Gamma_V e^{-2ikL} e^{ikx}\right]\right)$$



The expression inside the parentheses is the sum of two phasors, one of length $V_S/2$ (proportional to the right-going wave) and one of length $|\Gamma_V| V_S/2$ (proportional to the left-going wave). This total phasor, represented by the dashed vector, rotates counterclockwise at ω. Its projection on the x-axis is the physical voltage on the line. To find the sum, write the reflection coefficient as $\Gamma_V = |\Gamma_V| e^{i\theta}$. The phase difference between the two solid phasors is,

$$\varphi = kx + \theta - 2kL - (-kx) = 2k(x - L) + \theta$$

Now use the law of cosines to find the length of V_{total} :

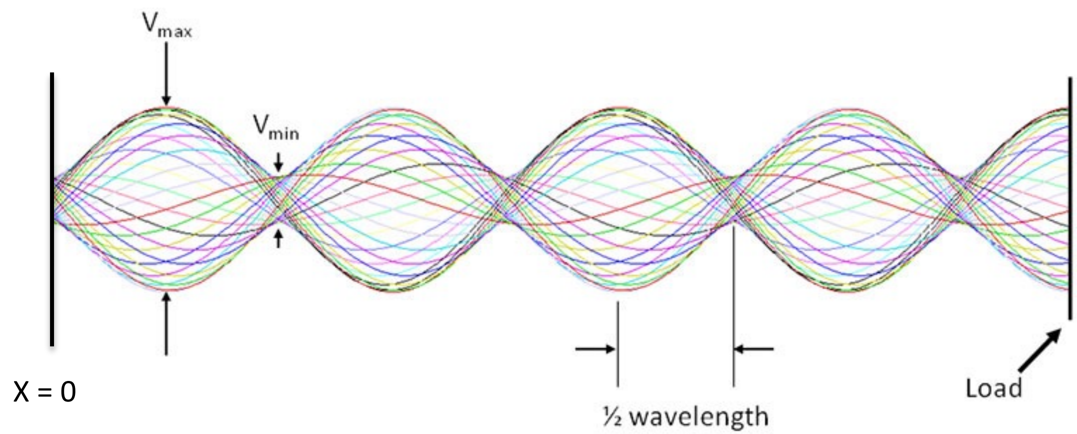
$$|V_{total}|^2 = \left|\frac{V_S}{2}\right|^2 + \left|\frac{V_S}{2}\Gamma_V\right|^2 + 2\left|\frac{V_S}{2}\right|^2 |\Gamma_V| \, cos\varphi = \left|\frac{V_S}{2}\right|^2 \left(1 + |\Gamma_V|^2 + 2|\Gamma_V| \, cos(2kx - 2kL + \theta)\right)$$

This quantity varies as we move along x because the cosine changes from -1 to + 1.

$$|V_{Total}|_{Max} = \left|\frac{V_S}{2}\right| (1 + 2|\Gamma_V| + |\Gamma_V|^2)^{\frac{1}{2}} = \left|\frac{V_S}{2}\right| (1 + |\Gamma_V|)$$

$$|V_{Total}|_{Min} = \left|\frac{V_S}{2}\right| (1 - 2|\Gamma_V| + |\Gamma_V|^2)^{\frac{1}{2}} = \left|\frac{V_S}{2}\right| (1 - |\Gamma_V|)$$

The standing waves have a minimum and maximum amplitude as shown below,



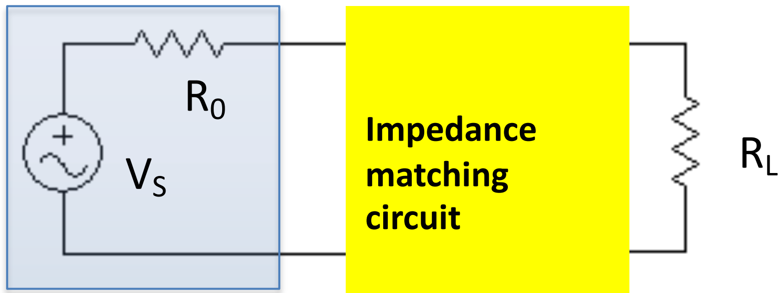
From: nutsvolts.com

The ratio of these two values is known as the voltage standing wave ratio, known as VSWR:

$$VSWR = \frac{V_{Max}}{V_{Min}} = \frac{(1 + |\Gamma_V|)}{(1 - |\Gamma_V|)}$$

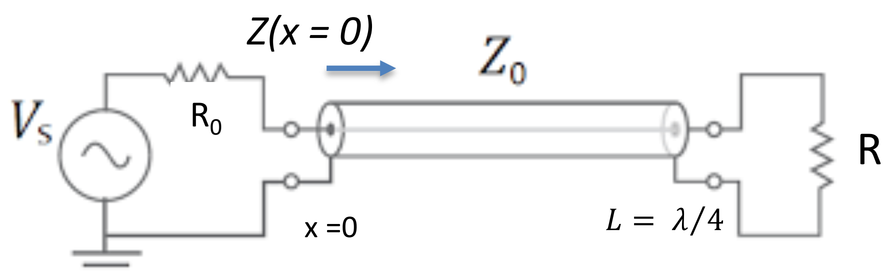
Impedance matching

Suppose we have a load resistance R_L to which we wish to transfer the maximum *power*. If the generator has the effective circuit shown in the shaded box, then the maximum power it can deliver to a load occurs when $R_0 = R_L$. However, R_L is generally *not* equal to R_0 so what can be done? The solution is an impedance-matching circuit that tricks the generator into thinking that it's connected to a load of resistance R_0. This can be done with L's and C's or you can use a transmission line to do it.



Once again, mexploit the impedance transformation properties of a $L = \lambda/4$ transmission line:

$$\tan kL = \tan\frac{\pi}{2} = \infty \quad \rightarrow \quad Z(x=0) = \frac{Z_0^2}{R_L}$$



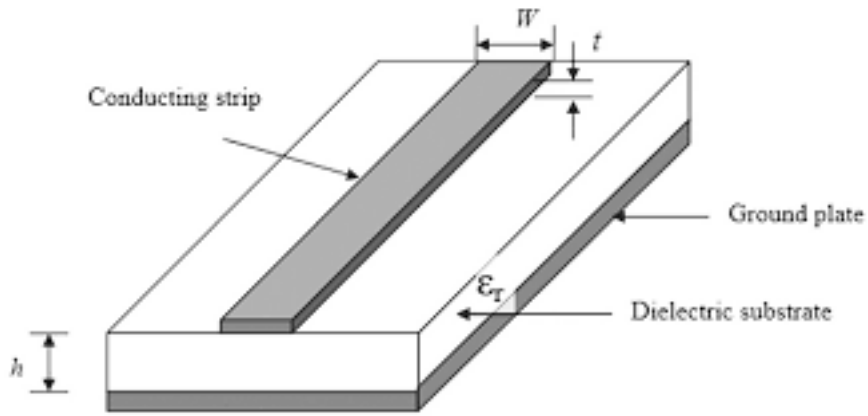
To impedance match R_L to the generator we choose,

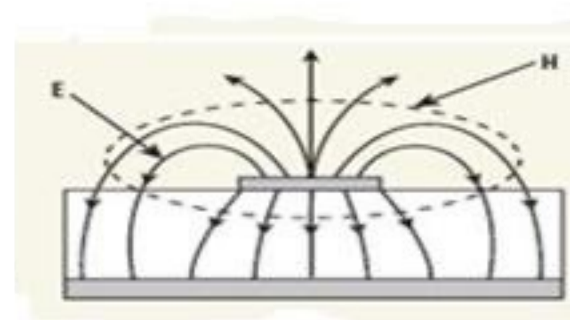
$$Z(x=0) = \frac{Z_0^2}{R_L} = R_0 \rightarrow Z_0 = \sqrt{R_L R_0}$$

So to impedance match the generator to the load, you need to make a transmission line with Z_0 equal to the geometric mean of R_0 and R_L. And it needs to be a quarter wavelength long at the frequency where you wish to operate. But how to make a transmission line with a specific characteristic impedance? The next slide shows one way.

Microstrip

Microstrip is a transmission line configuration that is widely used in high frequency circuits. It consists of a conducting strip sitting on a dielectric, beneath which is a conducting ground plane. The right-hand figure shows the configuration of **E** and **H** fields.





The nice thing about microstrip is that by adjusting the ratio *W/h* you can set the characteristic impedance Z_0 . An approximate formula is,

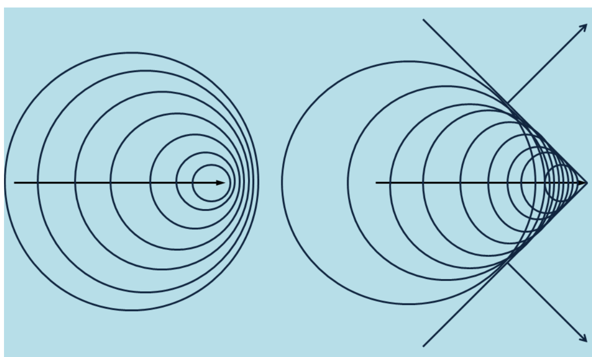
$$Z_0(Ohms) \approx \frac{377}{\left(\frac{W}{h}+1\right)\sqrt{\epsilon_r+\sqrt{\varepsilon_r}}}$$

As the frequency moves toward 1 GHz and beyond, this approximate formula no longer holds for several reasons:

(1) the dielectric constant is frequency-dependent and develops a resistive (i.e., loss) component.
(2) The transmission line model must include dielectric loss and a frequency-dependent resistance due to the skin depth of the metal.
(3) The field configuration shown is a transverse electromagnetic (TEM) wave in which E and H are perpendicular to the direction of propagation. This configuration holds true so long as W and h are much less than the wavelength of the wave. This, in turn, is typically valid for frequencies well below 10-15 GHz, after which non-TEM modes may appear and complicate things considerably.

Cherenkov radiation

Cherenkov radiation was first observed by Marie Curie more than a century ago and later explained by Cherenkov, Frank and Tamm. It involves the electromagnetic radiation by a medium as a charged particle passes through it very rapidly. The kinematic features are like the sonic boom from an airplane travelling faster than the speed of sound. In the figure below, imagine an airplane is travelling at a speed $v < c_S$, the speed of sound. Then sound waves emitted from the tip lead to spherical wavefronts expanding out from successive locations of airplane. None of the wavefronts cross each other so there is no interference between them. (An observer standing in front of the oncoming plane will hear a higher frequency than one to the left. That's the familiar Doppler effect. But now look at the right-hand figure. Since $v > c_S$ the plane moves *faster* than the wavefronts. Then the expanding wavefronts *do* cross each other so there is constructive interference. This leads to a wedge-shaped front which is a sonic boom.

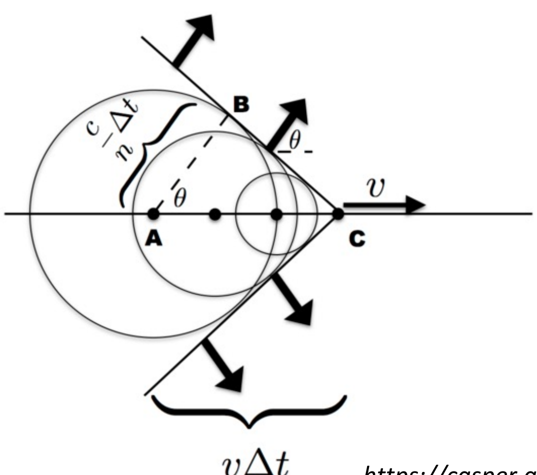


https://www.thphys.uni-heidelberg.de/~wolschin/eds14_3s.pdf

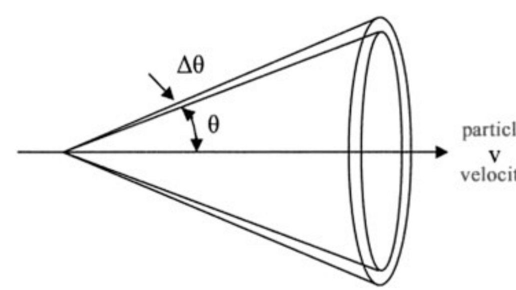
Now replace the airplane by a charged particle and the sound waves by light waves travelling through a material whose index of refraction is n. The phase velocity of light in the material is $c/n < c$. Think of the particle's electric field exciting atomic dipoles in the material as it moves along. These excited dipoles radiate out the waves shown. If $v > c/n$ then waves emitted from successive locations of the particle can interfere constructively and the resulting "shockwave" is Cherenkov radiation.

The lower figure shows the geometry. In a time Δt, light emitted at point A travels out to point B. By that time the particle has moved a distance $v\,\Delta t$ to point C. Dipoles at C are just beginning to emit radiation. So the shock wavefront extends along the line starting at C and extending back through B, defining a cone of radiation with an angle θ satisfying,

$$\cos\theta = \frac{c/n}{v}$$

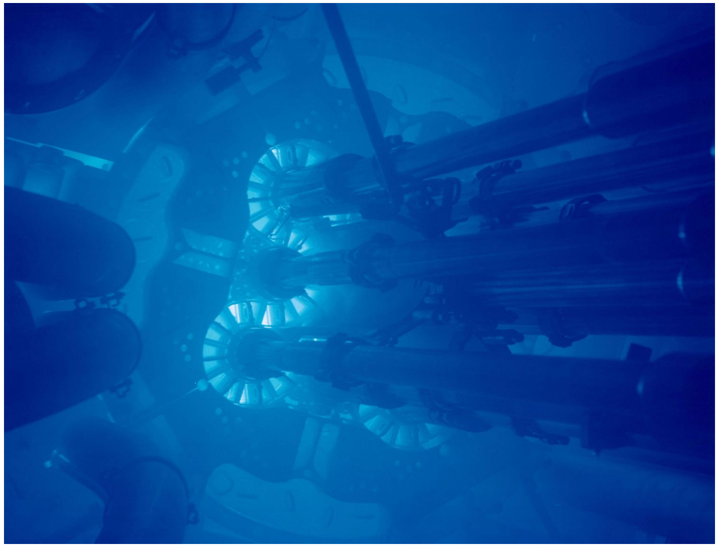


The Cherenkov wavefronts propagate at θ relative to the direction of the particle, defining a *cone* of radiation. (The thick arrows normal to the shock wavefront define the direction of emitted photons.) For the Cerenkov process to occur we must have $v > c/n$ and this in turn sets a lower limit for the energy of the charged particle. We'll leave it as an exercise to show that an electron moving through water (n = 1.333) must have a minimum kinetic energy of 263 keV in order to emit Cerenkov radiation.



https://casper.astro.berkeley.edu/astrobaki/index.php/Cherenkov Radiation

Cherenkov radiation requires charged particles moving at close to the speed of light. For example, in a nuclear reactor the fuel rods are constantly undergoing nuclear disintegrations that emit high energy electrons (β particles). The rods are surrounded by cooling water so the electrons generate Cherenkov radiation whose predominant visible component is blue.

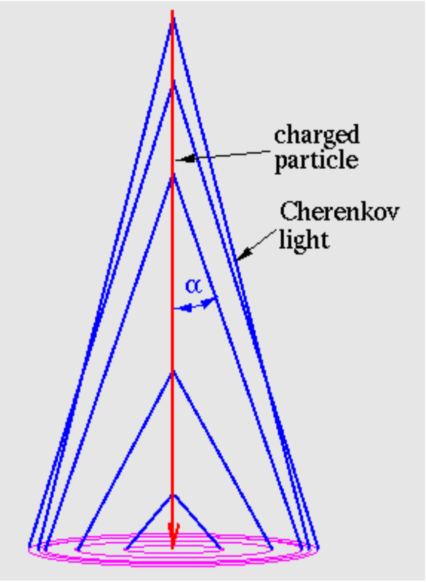


Why blue? We need to look more closely at the theory, which which is spelled out in, for example, J.D. Jackson, *Classical Electrodynamics*. The power radiated in a small frequency range from ω to $\omega + d\omega$ is given by

$$dP(\omega) = v\left(\frac{q}{c}\right)^2\left(1 - \frac{1}{\beta^2 n(\omega)^2}\right)\omega\, d\omega \qquad v > \frac{c}{n(\omega)}$$

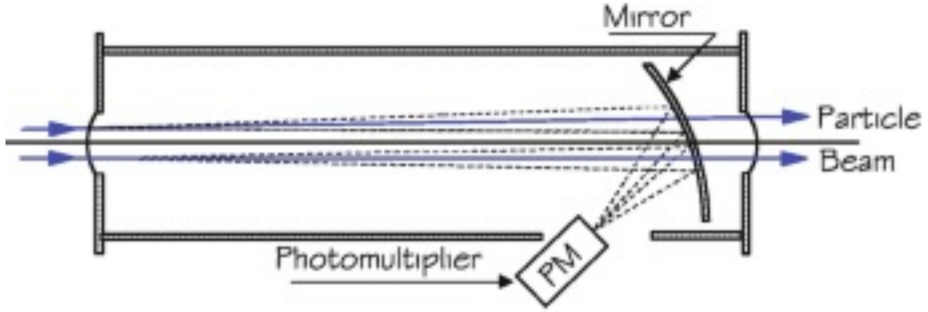
where *q* is the charge of the moving particle. In general, the index of refraction *n* depends on frequency so there can be frequencies where $\mathrm{v} < \mathrm{c}/n(\omega)$ (no radiation) and frequencies where $\mathrm{v} > \mathrm{c}/n(\omega)$ (radiation). The frequency dependence of the emitted power leads to more radiation at higher frequencies, so the visible component of Cherenkov radiation is blue.

Another example is atmospheric Cherenkov radiation, caused by cosmic rays. These are very high energy particles that enter the upper atmosphere, break up atoms and generate cascades of charged particles moving at very high speeds. These particles then generate Cherenkov radiation as they pass through the atmosphere. Very high up the air is thin so the index of refraction *n* is close to 1 and the angle of the Cherenkov cone is small. Closer to the earth the atmosphere is denser, *n* is larger so the Cerenkov angle is larger. Cones from radiation generating at varying altitudes overlap as shown in the figure. Atmospheric Cherenkov light was first identified in the 1940's. More recently, attention has been focused on detecting very high energy gamma ray photons that travel in a straight line from some interesting object out in space. These photons have energies exceeding 0.1 TeV and are rare but they produce charged particle cascades that in turn generate Cherenkov radiation. By detecting the occasional bursts of Cherenkov radiation with an array of telescopes it's possible to distinguish these special gamma rays from the general background of atmospheric Cherenkov radiation.



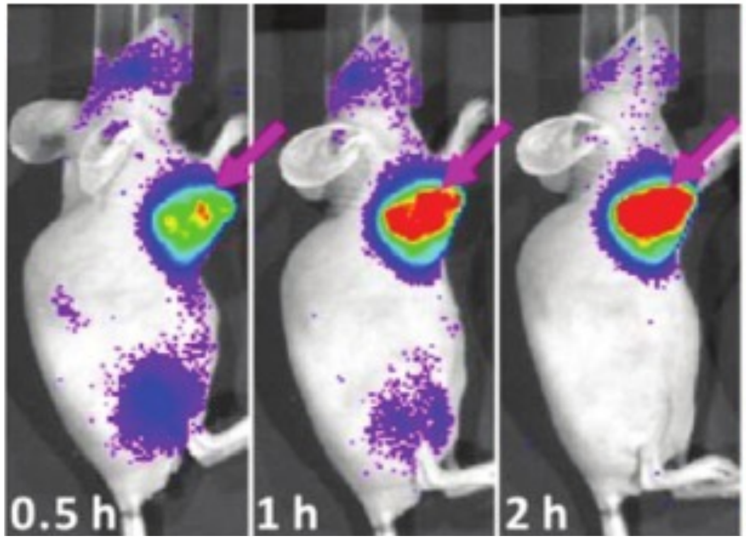
https://www.mpi-hd.mpg.de/hfm/CosmicRay/ChLight/Cherenkov.html

Particle physics makes extensive use of Cherenkov radiation to measure the speed of particles emitted from high energy collisions. The detector shown below gives the general idea. Fast particles enter a chamber filled with some dielectric, often a gas. The dashed lines indicate the Cerenkov cones around each particle beam direction. Cherenkov photons are reflected from the mirror and captured by the photomultiplier. Cherenkov photons generated by slower particles are emitted with *larger* angles than those shown and would not make it to the photomultiplier to be collected. The device therefore functions as a *threshold* detector, determining if the particle velocity is above or below some specified value determined by the geometry and the index of refraction of the medium.



https://link.springer.com/chapter/10.1007/978-3-030-35318-6_7

Cherenkov radiation has also been used to image cancer cells. The tumor is injected with radioactive nuclei which decay and emit very fast electrons or positrons which then move through the surrounding medium (mostly water) and generate Cherenkov radiation. The image on the left is from a mouse whose tumor was injected with radioactive ^{18}F.



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3477724/