CRYSTAL SETS TO SIDEBAND

Chapter 4
HERTZIAN WAVES IN THE BASEMENT

Crystal set radios and ancient spark gap transmitters from the 19th century may seem too simple and too far removed from ham radio to be worth building. If you're already familiar with basic electronics and early radio history, you may not learn anything from this exercise. If you're dying to build a real ham rig, please skip ahead to the next chapter. On the other hand, if you have little electronics experience, there are worlds of lessons to be learned from old technology. And if you've never built simple radios before, you should find it fun.

The nature of radio waves

Before we build transmitters and receivers, let's review radio waves. When we understand what radio waves are, the technology to generate and receive them becomes more obvious. An electromagnetic wave is an oscillation in free space that radiates out away from its source at the speed of light. It is called ***electromagnetic*** because it is both electric field energy and magnetic field energy. The wave oscillates or changes back and forth between these two forms of energy as it travels.

Propagating across vacuum

From our experiences with magnets and static electricity, it's hard to visualize how a magnetic or electric field can travel millions of miles across the vacuum of space. In our experience these fields are tightly localized around the device that generated them. How can a magnetic field exist isolated in a vacuum, perhaps even light years away from the nearest atom? Suppose we could somehow magically generate a magnetic or an electric field in space, miles from the nearest object. Would the field just sit in space forever waiting for an object to pass by and be influenced by the field?

Let's suppose there is a refrigerator magnet floating in the void of space. Its magnet field will be at rest in the space surrounding the magnet, just as it does on your refrigerator. As always, the magnetic field will reach out its usual distance of a quarter inch or so. However, if the magnet were to suddenly vanish, the energy in the field would lose its "container" or "anchor" and be loose in the void.

The same scenario can be proposed for an electric field: If a flashlight battery were floating in space, the electric force would extend perhaps an inch out into the space in a halo around the two battery terminals. Again, if the battery suddenly vanished, the electric field energy would lose its generator and be stranded in the void. Without its anchor, it would spill out in all directions.

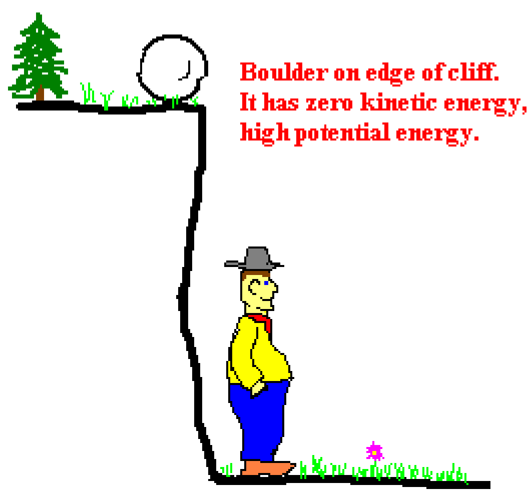
Oscillation occurs when two form of energy rhythmically change back and forth

When magnetic and electric fields are turned loose in space, what becomes of them? As James Maxwell first explained over a century ago, electric fields and magnetic fields are intimately related. It turns out that ***a changing or moving electric field generates a changing***

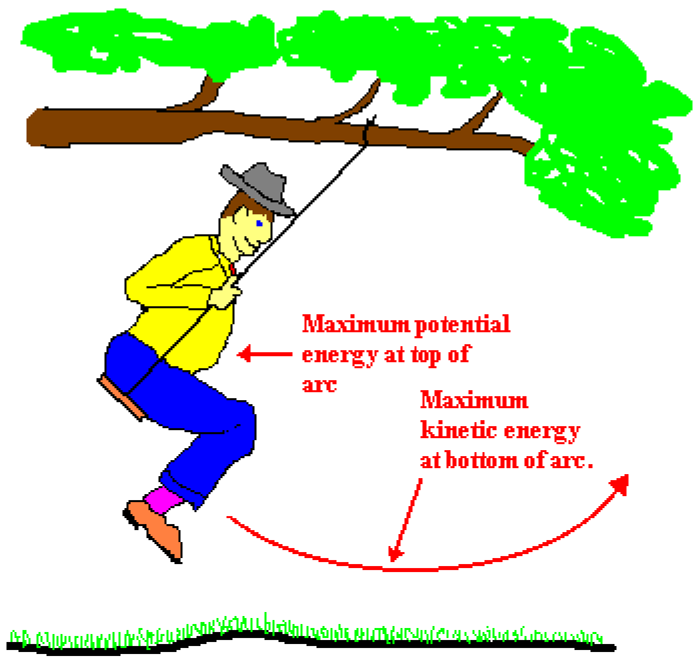
magnetic field and vice-versa. As the energy continues to "spill" out in all directions, the energy oscillates back and forth between these two kinds of fields. This connection is not "obvious" or intuitive. If it were, the Greeks, Chinese, or Egyptians would have described and exploited it long ago.

Mechanical oscillators

Many physical devices in our world oscillate, so oscillation between electric and magnetic fields should not be a surprise. An oscillation in nature can be described as energy spontaneously transforming from one form of energy into another then back again. For example, as a clock pendulum swings back and forth, the pendulum acquires the kinetic energy of motion as it swings through the bottom of its arc. Then, when the pendulum swings back uphill, the energy contained in the kinetic energy is returned to gravitational potential energy. When the pendulum reaches the top of its swing, it momentarily comes to a complete halt, turns around and races back downhill. When it is at the top, the energy is all "potential." A boulder sitting on the lip of a cliff doesn't seem to have any energy until it is nudged off the cliff. The fellow standing at the bottom of cliff can testify that the rock has plenty of energy when it slams into the foot of the cliff. (That assumes, of course, that he survives.)



To reiterate, an oscillating pendulum switches its energy back and forth between kinetic energy and potential energy. Notice that the length of the pendulum establishes the frequency of the oscillation of a pendulum. This is because gravity is constant and lightweight things fall just as fast as heavy things. If you ignore air resistance, the frequency of the pendulum swing is determined solely by the length of the pendulum arm (and of course the acceleration of gravity). This makes a pendulum good for keeping a mechanical clock running uniformly and accurately. Similarly, once the frequency of a radio wave is established, it doesn't change frequency as it races through space and becomes weaker.

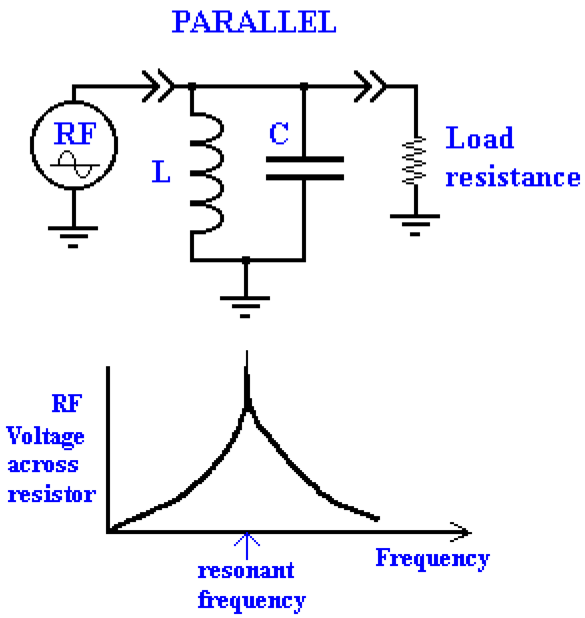


In summary, free space (which is literally "nothing"), can support magnetic or electric field energy, but only temporarily. To be maintained, a magnetic field needs to be generated by a device. ***A magnetic field can be generated temporally by a nearby collapsing electric field. As the magnetic field collapses, it produces a temporary electric field in the adjacent space.*** This seesaw produces a radio wave traveling outward across the void at the speed of light.

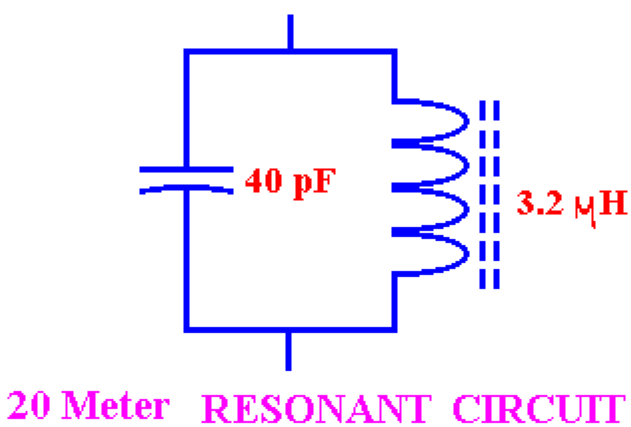
What's special about the speed of light? Good question! Einstein's theories of relativity teach us that time and space are frozen like a 4 dimensional ice cube. Most people are aware that light speed is the ultimate velocity. What is rarely understood is that light speed is really "the speed of time itself." Relativity is rarely studied deeply, even by engineers. This is perhaps because, until the invention of portable atomic clocks, the evidence proving relativity was less than overwhelming. For example, when the first GPS satellite was launched, the software had "switches" that allowed ground engineers to turn off the adjustments for Special and General Relativity in case relativity turned out to be wrong. They weren't. Our universe is weird and profound far beyond what 99.9% of humans have the privilege to notice. However, these subjects belong in another book.

Getting back to 20th century radio: Transmitter antennas are designed to generate either a rapidly changing electric field, or alternatively, a rapidly changing magnetic field. The antenna is placed out in the open with free access to the sky. The electric or magnetic fields around the antenna create the opposite kind of field and the result is a free-flying radio wave. The same antennas work well for receivers. As radio waves flash past the metal antenna elements, electric currents are induced into the structure just as if it were temporarily a capacitor or a secondary loop of a transformer.

The LC circuit, the fundamental electronic oscillator



The most fundamental component of all radio transmitters and receivers is the ***capacitor/ inductor parallel resonant circuit***. This basic circuit consists of an inductor wired in parallel with a capacitor. These are called ***LC circuits*** where L is the letter used when calculating inductance and C of course stands for capacitance. If a high frequency sinewave voltage is applied across the parallel LC circuit, there is a specific frequency at which the LC circuit resonates and appears to be an open circuit. At all the other frequencies the LC appears as a load or short circuit. The LC circuit attenuates or eliminates the sinewave at every frequency except one. In this way one radio signal can be "tuned in" preferentially over another.

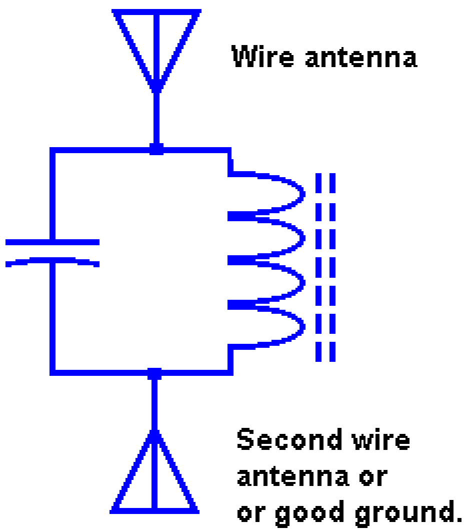


For example, the circuit above resonates at 14 MHz, the 20-meter hamband. The tiny inductor is just 3.2 microHenries. The capacitor is only 40 picoFarads which means 40 millionths-of-a-millionth of a Farad.

The LC circuit is a kind of electric oscillator. It is analogous to a swinging pendulum or a weight bouncing up and down on a mechanical spring. The LC oscillator goes through the same energy cycle as radio waves. First the energy is stored in the magnetic field inside and around the inductor. During the next half cycle, the energy is stored in an electric field between the plates of the capacitor. The energy alternates back and forth between these components until the resistance in the wires dissipates it.

A parallel resonant LC circuit with dual antennas forms a simple transmitter

Significant energy can also be leaked off into the space around the LC circuit as radio waves. Therefore, once we get an LC circuit to oscillate, we are already on our way to generating radio waves. ***If we simply add wires to the ends of the parallel LC circuit, these wires form an antenna to couple the electric field to the free space around it.*** In other words, the simple circuit shown below is a crude radio transmitter.

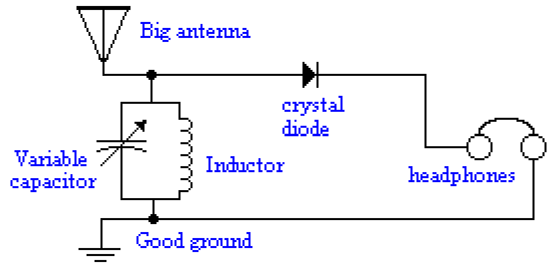


A receiver of radio waves can be built in the same way. Imagine that an identical parallel LC circuit with antennas is tuned to resonate at the same frequency. Now imagine that this second circuit is floating in the void, perhaps miles from the transmitting circuit. When the radio waves radiate past the receiving LC circuit, the electric field component in the radio wave will produce a tiny surge of current in the wires that charge the capacitor. Alternatively, and depending on the orientation of the coil with respect to the radio waves, the magnetic component of the radio wave will induce a tiny voltage to appear across the coil. This is the same as if the inductor were the secondary of a transformer. Once the radio wave has zoomed past, a tiny, miniscule oscillation will remain in the receiving LC circuit, ringing back and forth between the inductor and capacitor.

****

CRYSTAL SET RADIOS

A crystal set radio is an excellent first radio project. They have few parts and are easy to understand and build. Crystal sets were a common toy when I was a kid and my first electronic project was building one. Although we kids played with crystal sets, we didn't really understand them. When they didn't work, we had only the haziest notion of how to fix them. If you bought a commercial toy crystal set, I never saw one that had explanations even remotely as complete as what you are now reading. Starting with the crystal set, later on we can build on what we learn to build transmitters and more elaborate receivers.



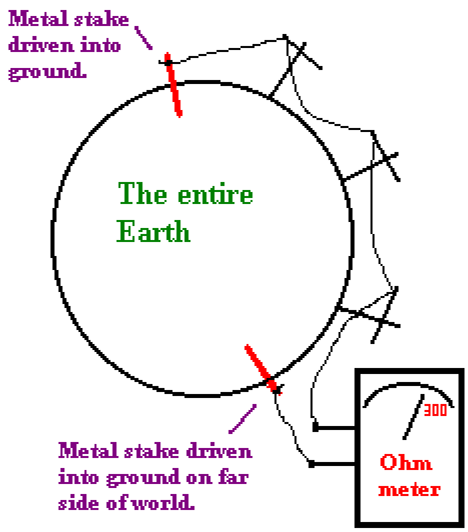
The combination of inductor and capacitor tune in the desired station. That is, the inductor and capacitor oscillate at the frequency of the desired station. The crystal diode rectifies the sinewaves oscillating across the LC circuit. This converts the high frequency sinewaves into low frequency sound frequencies that can be heard in the headphones. Crystal sets have six basic parts. The antenna of course picks up the signal from the air.

The antenna

For crystal sets it is easiest to build an "electric field" antenna. It is usually just a long piece of wire strung out a window or up in a tree. A limitation of a single wire antenna like this is that, when the radio wave generates a voltage on the wire, the current it might produce has no where to go. A simple wire antenna is like one terminal of battery. Yes, the battery has a voltage, but without a connection to the other terminal of the LC circuit, the current has no circuit to flow through. To provide a destination for the current we can add a second antenna. Alternatively we can connect the crystal set to "ground."

The ground

Electrical "ground" is a word that we learn as youngsters, but most people go through their whole lives without ever understanding it. I suspect that the term arose during the early days of telegraph communication in the 1840s. It turns out that wet earth is a fairly good conductor. If you drive two metal stakes into the ground in your backyard and connect a battery to the two stakes, current will flow from one stake to the other. For two stakes about 100 feet apart, the ground has a resistance of about 100 ohms. If you experiment, you learn that most of the electrical resistance to current flow occurs right around the stakes. Once the current gets launched, the electrical resistance only increases slightly as you increase the distance. A metal stake in a backyard in China and the one in your backyard in the US might only have 300 ohms resistance between them. This was a boon for early telegraphers because it meant they only had to string one wire between cities instead of two wires to complete the circuit loop. In practice, using ground as an intercity "wire" is not as reliable as stringing a second wire, but it illustrates the concept of ground.

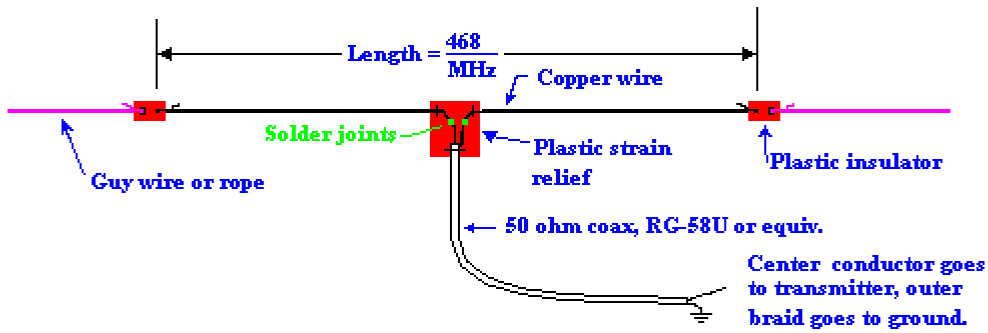


A good electrical connection with ground is an essential component of the crystal set. The most accessible ground for a crystal set or a ham radio station is usually a copper water pipe or a hot water heating pipe. At low frequencies, like standard broadcast AM, ideal electric antennas are very large. Needless to say, being able to use the entire Earth as half of the antenna is often quite convenient.

Dipoles

Unlike low frequencies, at high frequencies, like VHF television or FM radio, the ideal length for an electric antenna is just a meter or so. The higher the frequency, the less distance there is between successive electric and magnetic waves. That is, wavelength decreases with increasing frequency.

Although in theory you could use a stake driven into your lawn for the "ground" path on your TV, it is much simpler to just use a second short antenna oriented 180^{o} away from the "real" antenna. This dual antenna is called a ***dipole*** and is the most common basic antenna design used in ham radio and TV antennas. Although it isn't obvious, the arrays of thin metal tubes on rooftops are just refinements of the basic dipole antenna. A common and versatile ham antenna is the dipole shown below. In general, the lower the frequency, the larger the dipole must be in order to work well. A typical ham radio dipole is shown below.

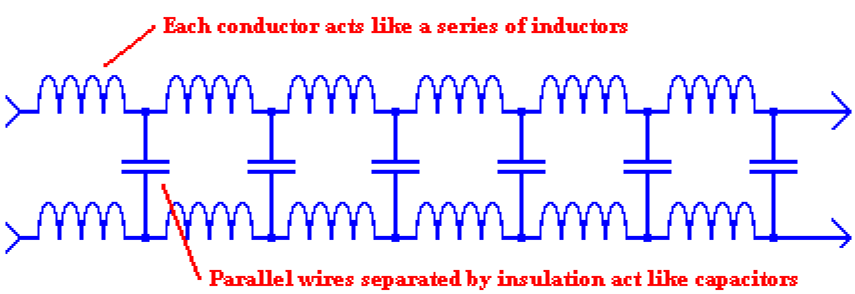


For 40 meters the length is 66 feet. The coax feed line can be any length, but shorter is better.

Transmission lines

In the ham radio dipole antenna above, the "arms" stick out in space in opposite directions and snag the passing electrical field. In general, the higher the dipole is above the local terrain, the better the dipole will receive signals. Unfortunately, you and your radio are down on the ground. Climbing up on the roof to listen to listen to the radio or watch TV is inconvenient, to say the least. The problem of how to move radio frequency signals down to the receiver is solved with a ***transmission line.***

A transmission line is a pair of parallel wires separated by insulation. It works very much like a speaking tube in a ship or even the string in a tin can telephone. In all these devices, vibrations are transmitted down a narrow pathway with surprisingly little loss of energy. A terrific example of a mechanical transmission line can be a farmer's ditch filled with water. Provided the water is flowing slowly, when you throw a big rock into the ditch, the wave from the splash will travel hundreds of yards before it dissipates. A wave in the ditch propagates unchanged for many minutes and travels great distances. In contrast, if you throw the same rock into an open pond, the wave spreads out in all directions and quickly vanishes.

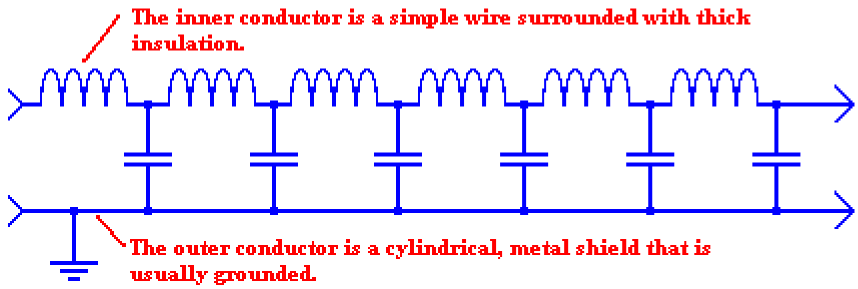


Two parallel wires separated by insulation make up a transmisison line.

A radio transmission line is a distributed, LC resonant circuit. We saw in the chapter 2 that a simple wire has inductance. Similarly any two wires separated by an insulator comprise a capacitor, whether we planned on making a capacitor or not. Consequently, when we run two parallel, insulated wires over any distance, there will be a measurable capacitance between them and the wires themselves will have a significant amount of inductance. To a radio wave, this construction looks like a long, LC circuit without end. As the capacitance and inductance are charged and discharged, the oscillation doesn't stand still, but rather moves down the pair of wires at nearly the speed of light. As you can see, propagation down a transmission line is analogous to propagation through free space, but it only propagates in one dimension instead of three dimensions. The electric field or voltage generates a current and magnetic field, which in turn generates a new electric field and so on. An example of a simple transmission line consisting of two parallel wires is the flat wire "300 ohm" TV line used to feed older TVs.

Coaxial cable

The round, shielded transmission line used to feed modern TVs is a ***coaxial cable***. Instead of using two separate ordinary wires, the outer conductor of a coaxial cable is a metal cylinder that completely encloses the center conductor. The inductance of the shield conductor is far less than that of a simple wire, but it prevents the radio frequency signals on the inner conductor from leaking out. Even better, it keeps new signals from leaking into the cable and interfering with TV reception.



A coaxial cable is a transmission line with a cylindrical, outer conductor shield.

The ham radio dipole antenna described earlier uses type RG-58 coaxial cable to transport the radio frequency signals down into the house. You will notice that the outer shield is connected to ground. This is nearly always the case with coax. You could use cheap TV coax cable for your ham transmitter, but you will find it difficult to work with. The outer shield of cheap coax is just aluminum foil and is difficult to connect mechanically and electrically. In contrast, the outer shield of quality coax is braided copper wire that is easy to cut and solder. It also has considerable mechanical strength.

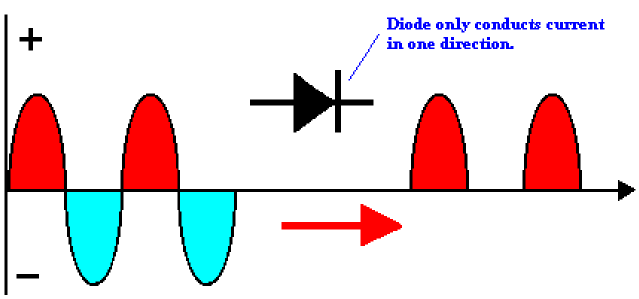
Transmission line impedance

An abstract characteristic of transmission lines is that, from the point of view of a traveling radio signal, the line "looks like" a specific load resistance. For example, RG-58 coax appears to the radio signal to be a 50 ohm resistor. It isn't, of course, but the voltage and current levels along the wire suggest that it is. In other words, voltage divided by current at points along the line will give 50 ohms. Another reason not to use TV cable is that TV cable is usually designed for 75 ohms, while most ham equipment is designed for 50 ohms.

In general, the finer the wire and farther apart the two conductors of a transmission line, the higher the characteristic impedance. The flat, brown, ribbon transmission line that was formerly in common use for TV antennas has an impedance of 300 ohms. Sometimes hams use a wide "ladder line" in which bare copper conductors are separated by an inch or more of air and a few ceramic separators. Ladder line often has an impedance of 600 ohms. Ladder line is useful when transmitter power must be transmitted great distances to get to the antenna. Because a ladder line has little or no insulation in contact with the wires, the small dissipation of energy in the insulation is reduced to the absolute minimum. More importantly, the high impedance means that smaller currents will flow in the wire and less energy will be dissipated in the resistance of the copper. Of course, the trade off is that higher voltages are needed to transmit the same power levels. I know a ham who lives at the bottom of canyon where radio reception is poor. He uses a long ladder line hung in the trees to connect his transmitter to the antenna, which is located hundreds of yards up on the mountainside.

Diode detectors

The diode is the ***detector*** that converts radio frequency sinewaves into audio frequency electric waves, ready to be converted into sound. The diode is a "one way electricity valve." In plumbing terms, it works like a check valve. The schematic symbol for a diode is an arrowhead pointed at a barrier at right angles to the wire.



HALF-WAVE RECTIFICATION

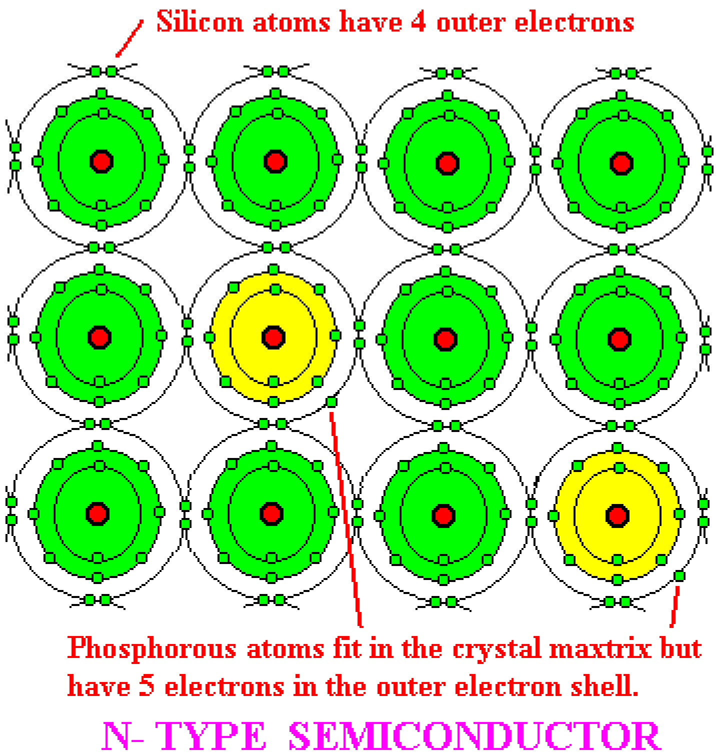
(Only one half of wave is exploited.)

In electronics, the convention for "***positive" current flow is from positive to negative***. Unfortunately, the flow of electrons is from negative to positive. So, what is actually "flowing" from positive to negative is the absence of electrons. Confusing, no? I suspect this convention was established before electrons were understood. Referring to the symbol for a diode, positive current is allowed to pass if it flows in the direction of the arrowhead. Positive current will be blocked by the diode if it attempts to enter the diode from the perpendicular "barrier" side.

Semiconductors usually perform the check-valve function of diodes. A semiconductor is a crystal of an element like silicon or germanium that has a chemical valence of 4. That is, during chemical reactions this element can either take up 4 electrons, or give away four electrons. As we shall see shortly, semiconductors can also be fashioned by making crystals out of mixed elements with valences of 3 and 5, or even 2 and 6.

N-type semiconductors

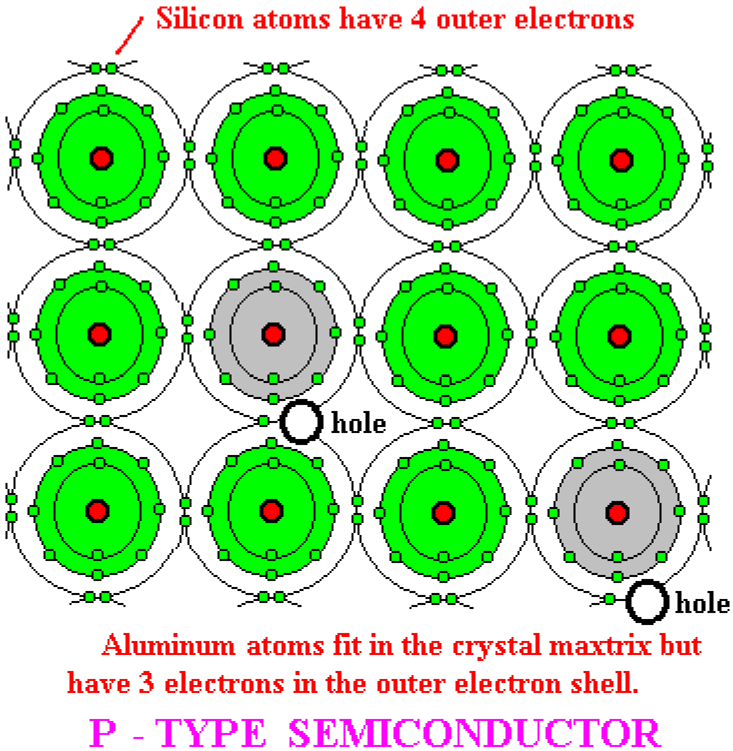
To keep it simple, let's suppose we have a pure crystal made out of silicon, which has a valence of 4. If we put some multimeter (ohm meter) probes across this pure silicon, it will act like an insulator – there will be no significant current flow. However, if we make a new crystal with just a touch of phosphorus impurity in it, suddenly it becomes a conductor. Phosphorus has a valence 5 and is almost the same as silicon in atomic weight. This means that in chemical reactions it normally accepts 3 electrons to complete an outer shell of 8 electrons. But when silicon crystal is contaminated with phosphorus, lone atoms of phosphorus are trapped among a frozen, rigid crystal of silicon. The phosphorus atom fits in the matrix, but it has an extra electron that is "loose" and free to move around the crystal. The electron can't move over to silicon atoms because they are joined with neighboring silicon atoms so that each silicon atom has a stable outer shell of eight shared electrons. However, the extra phosphorous electron can move over to other phosphorous atoms that have already lost their 5th electron. In other words, ***a silicon crystal with just a touch of valence 5 impurity acts like metal.*** It has electrons that are free to migrate through the whole solid. A semiconductor with extra electrons is called an ***N-type semiconductor.***



P-type semiconductor

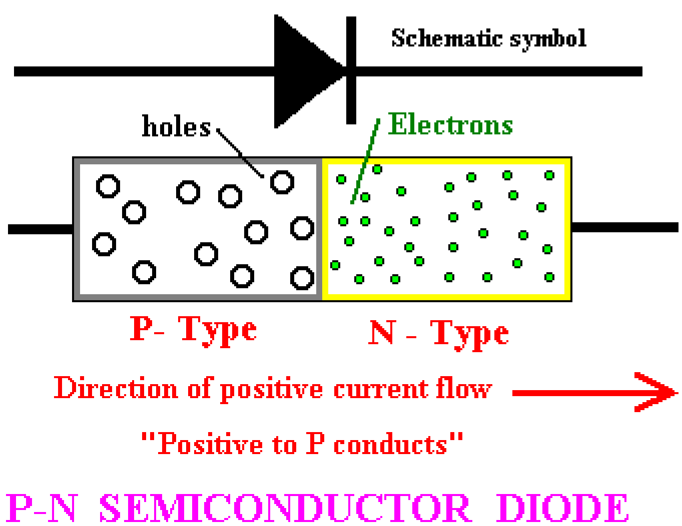
P-type semiconductor is a bit abstract. Instead of making a silicon crystal with valence 5 impurity, now suppose we add an impurity such as aluminum, indium, or gallium with a valence of 3. The impurity fits into the crystal matrix, but it needs one more electron to reach an equilibrium of 8 electrons shared with its neighboring silicon atoms. In other words, this semiconductor has "***holes***" in the crystal matrix that can be filled by electrons passing through. Now when you place multimeter probes across a P-semiconductor, it will conduct just like the N-type semiconductor. However, the conduction mechanism is different. With P-type semiconductor, the negative metal probe touching the crystal supplies all the free electrons flowing through the crystal. These electrons are hopping from hole to hole to cross the crystal.

Like phosphorus, aluminum atoms have almost the same atomic weight and size as silicon. Aluminum atoms fit perfectly in the silicon crystal matrix.



Diodes are P-N junctions

Semiconductor diodes are constructed by placing P- type semiconductor in contact with N-type semiconductor. In other words, for electrons to flow through the diode, the electrons must enter the N-type crystal and then move across the junction into the P-type where they complete the journey by jumping from hole to hole.



"positive to P conducts"

If we measure the resistance across a diode with an ohmmeter, the resistance is low when we place the positive pole of the meter on the P-type semiconductor and the negative pole on the N-type. OK. Now let's reverse the probes of the ohmmeter. We are placing the positive probe against the N-type semi-conductor and the negative probe against the P-type. Electrons flow off the metal probe and into the P-type semiconductor. No problem so far. On the other side of the diode the extra electrons from the N-type silicon are being attracted or "sucked" into the positive metal probe. Thus the conduction seems to start out all right, but it isn't long before the extra electrons in the N-type silicon along the P-N junction are depleted. All that remains in this region are depleted valence 5 atoms that are now acting like pure silicon. This whole region now acts like pure silicon and the conduction stops.

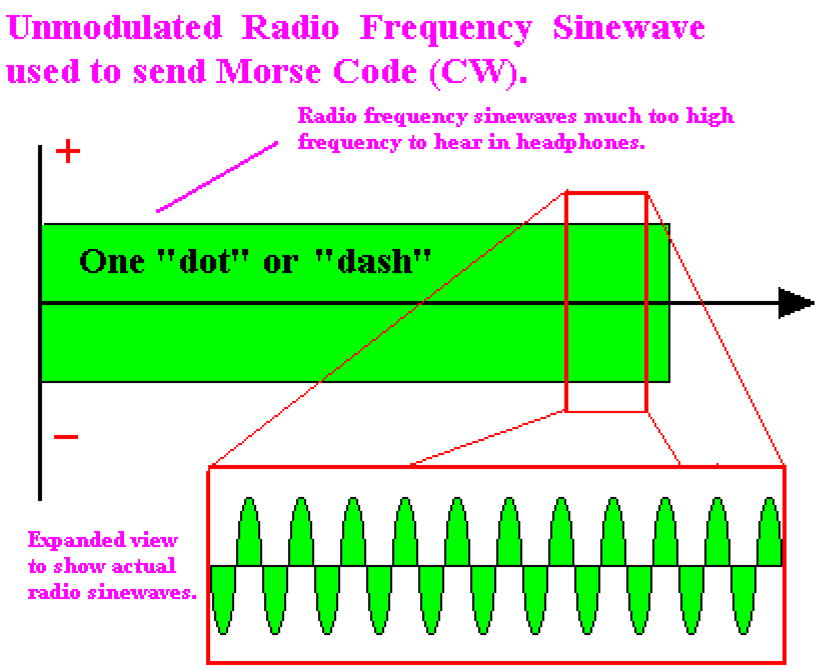
So why can't the electrons that are migrating through the P-semiconductor holes hop across the P-N barrier and move onto the valence 5 atoms? The reason is the same. The migrating electrons have filled in all the holes in the P-type and the crystal has also become pseudo-pure silicon that is an insulator. When thinking about PN diodes, remember, ***"positive to P conducts."***

Real diodes

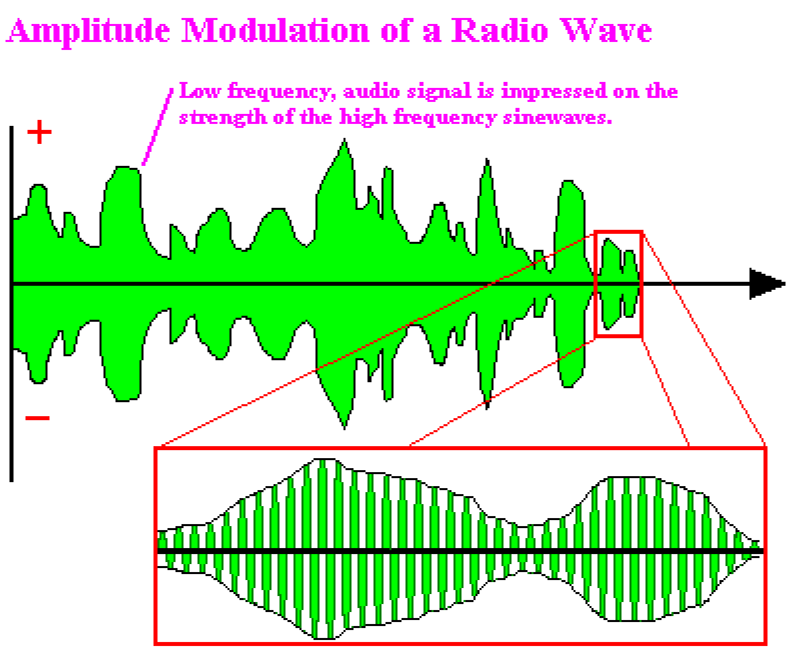
Commercial diodes come in all sizes. The types suitable for detectors in crystal sets are the little bitty guys on the lower left. The big black diodes on the right are rectifier diodes and diode "bridge" arrays for power supplies. The long black diode is rated at 6,000 volts. The two diodes that are built like machine bolts are high speed, high power units that might be used in a large switching power supply or an industrial RF generator.

Detection of AM radio signals with a diode

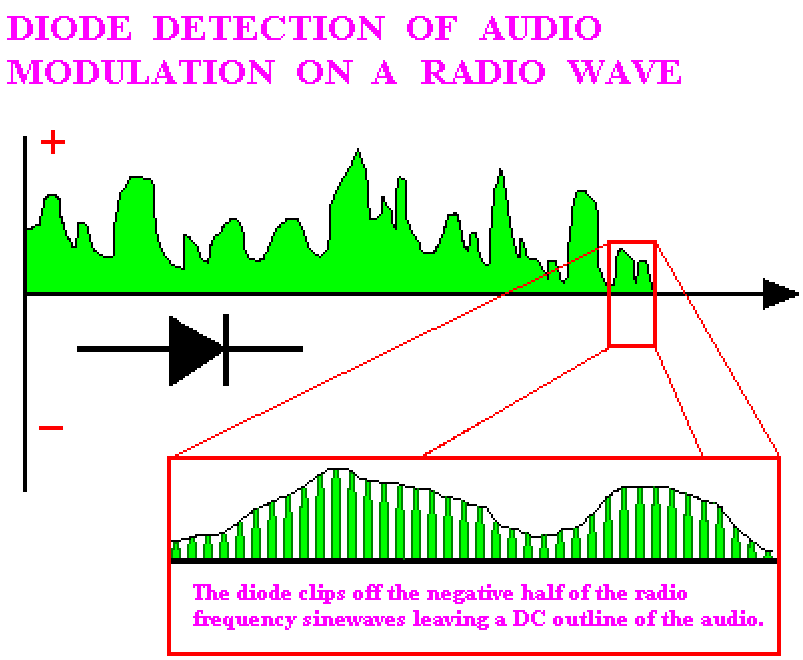
In amplitude modulation, (AM) the audio speech signal is impressed onto the radio signal by varying the AMPLITUDE of the radio signal. An AM transmitter literally increases and decreases the output power of the transmitter in time with the speech and music being broadcast. The drawing below shows an unmodulated radio signal of the sort used to send Morse code. The radio frequency sinewave remains the same amplitude throughout the time that the transmitter is keyed. Because the sinewave maintains its amplitude during the "dots" and "dashes," Morse code signals are known as ***continuous wave*** or "***CW***."



As the name implies, an ***amplitude modulated (AM) radio broadcast, (550 KHz to 1.750 MHz)*** makes a continuously varying graph of the RF signal that looks like a psychiatrist's Rorschach. But of course the outline of the audio signal is actually made up of hundreds of thousands or millions of RF sinewave cycles.



The diode detector recovers the audio signal by "shaving off" one of the two polarities of the RF signal. Sinewave currents have both positive and negative polarity. Diodes only allow conduction in one direction. So, when a radio frequency sinewave current is passed through a diode, one of those polarities will not pass and will be eliminated. What remains is a series of narrow, direct current pulses, all with the same polarity.



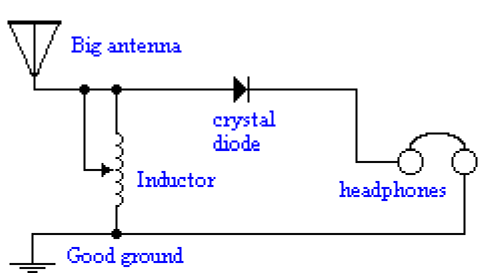
This detection process, which is also called ***rectification***, produces a varying DC signal that may be passed through a headphone to convert it into sound. Physically, a modern diode is usually a tiny glass cylinder typically ¼ inch long with two wires extending from the ends. It's not much to look at. As will be described below, it's more fun to make a diode out of sulfide ore, or even out of razor blades or safety pins.

Headphones

After the diode has generated the varying DC current representing the audio signal, a device is needed to convert the current into sound. The classic way to do this is to use a magnetic headphone. As we shall describe below, a headphone is an electromagnet that attracts a thin, steel diaphragm and makes it vibrate in time with the speech and music.

A practical crystal set schematic

The crystal set can be extremely simple. A schematic is shown below:



Crystal Set Parts List:

Big antenna – 50 feet of wire strung up in a tree will be ideal. Or, use the 40 meter dipole described above. For this application, use the entire dipole assembly as if it were a single piece of wire. Connect the center conductor and the braided outer shield of the coaxial cable together and fasten the resulting "wire" to the "big antenna" location above.

Good ground – A connection firmly clamped to a household copper water pipe would be ideal. Alternately, you may use a second length of wire strung up in another tree. The second wire should be far away from the first wire. I happen to have a 30 meter ham band dipole in my back yard. I used my 40 meter dipole as the "antenna" and the 30 meter as a "ground." Or, as an antenna would be known in this application, the 30 meter dipole became a ***"counterpoise."***

Inductor Wind about 20 turns of bare copper wire around a large diameter cardboard tube. Cardboard Quaker Oats boxes are the classic coil form for this purpose, but the cardboard cylinder from a toilet roll will work. In general, the larger the diameter coil, the better it works. I believe large coils work better because the coil is acting as a magnetic antenna, as well as a tuned LC circuit. In other words, a large diameter coil snags more magnetic field component from the radio wave. To tune the crystal set, you need to rig up a slider or shorting clip that allows you to short out some of the coil.

Capacitor Where's the variable capacitor? A capacitor consists of two pieces of metal separated by an insulator. If you wind a big coil of wire around a cardboard tube, then there is capacitance between one loop of wire and all the neighboring loops. "But, hold on! That can't be! They're shorted together!" you say. Yes, you're right. But if you look at an LC circuit as a whole, the inductor is a kind of "short circuit" across the whole capacitor and we know that works OK. The hard part about physics is that you have to learn to think abstractly. A lot of phenomena seem fuzzy and inconsistent. We are forced to "get a feel" for what works and what doesn't. The coil of wire is said to have ***intra-winding capacitance*** that acts the same as if it were a separate capacitor across the whole thing, honest.

Crystal diode It's great fun to build your own diode as described below. However, to get started, you may want to use an ordinary, small silicon diode such as a 1N4148 or a 1N914, which is available at Radio Shack.

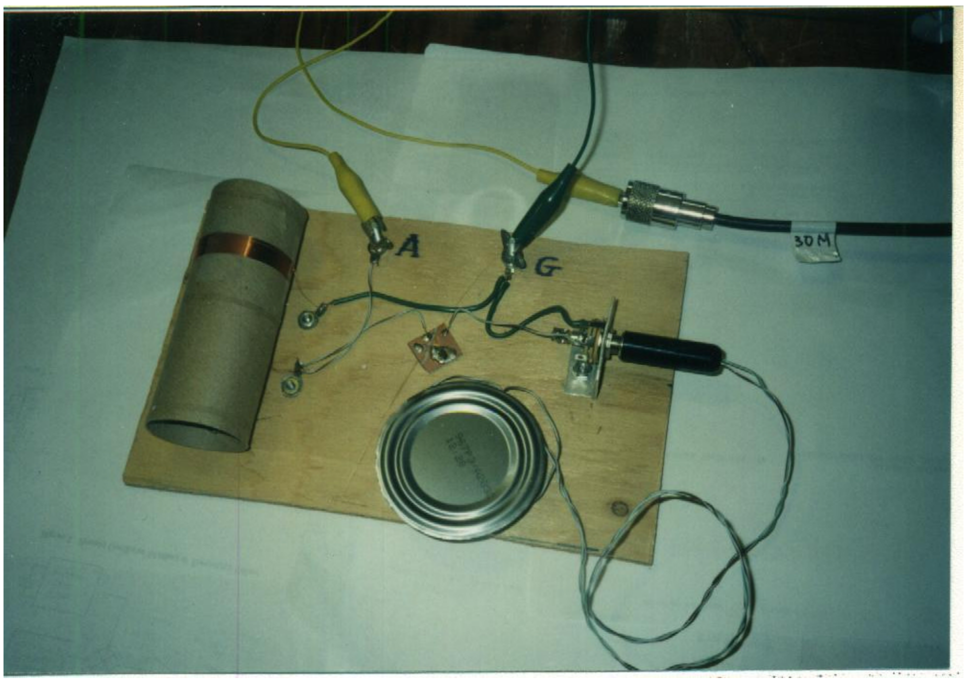
Headphones. You may construct a working headphone from ordinary parts as described below. This will be fun and educational, but sooner or later you will need to buy a good pair. You may buy either old-fashioned high impedance (2000 ohms) headphones or modern low impedance (8 ohms) headphones. The modern ones are extremely efficient, comfortable to wear and have hi-fi sound. The high impedance headphones are historic and little more can be said in their favor.

****

Homebuilt diode detectors

My experimentation with crystal sets as an adult began one day when I was hiking near Jamestown, Colorado. I was scrambling up a yellow-colored abandoned mine dump. Mine tailings up there are mostly yellow, sulfated, powdered rock that consists of broken-down granite or gneiss. Suddenly right in front of my face were chunks of the shiny, black sulfide ore that was the reason for the mine. Without an assay, I don't know exactly what's in this ore, but it's a safe bet that it's a mixture of sulfides of silver, lead, and maybe zinc, a dash of arsenic, tin, and copper. There might even be a trace of gold telluride in those crystals. Galena, which is lead sulfide, is the stuff used in old-time crystal sets to make detector diodes. “Gee! I wonder if I can make a crystal set out of this ore?”

It seems to me I once saw a war movie in which a POW in a Nazi Stalag made a radio out of barbed wire, a razor blade, and silver paper from a chewing gum wrapper. Well, that’s Hollywood, but maybe a receiver can be built without using parts specifically manufactured for radios. I happened to have a toy crystal set radio dating from about 1950 in my attic, so I hauled it down and checked it out. The “diode” consists of a tiny chunk of gray galena sticking out of a little puddle of solidified solder. The positive pole of the diode is a metal “cat whisker,” a piece of thin copper wire poked against the crystal.



The above picture shows my crystal set with all three homemade components. An LC resonant circuit is vital to select the AM radio band (or other band). An actual tuning capacitor isn’t really needed at AM radio frequencies. A big coil, at very least two inches in diameter with 20 to 60 turns wound on a cardboard tube, has enough inter-winding capacitance to resonate in the AM band. From a physics point of view, the circuit in the “practical” schematic is functionally the same as the circuit that contains the variable capacitor across the coil.

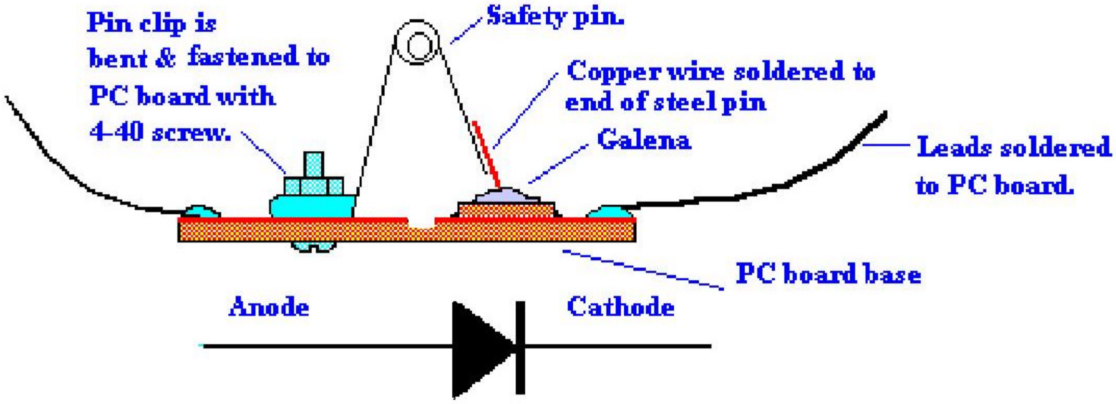
Without a variable capacitor, you will have no way to tune in particular stations. A tap on the winding can be added for peaking a station. A “tap” is just a way to short out part of the inductor. Using this method you can crudely, very crudely, select the loudest stations at the top or bottom of the AM broadcast band. However, if you prefer to use a variable capacitor, homemade or otherwise, be my guest. You will find that tuning a crystal set is (usually) sloppy no matter how you do it.

The crystal diode rectifies the radio frequency voltage ringing on the LC circuit and the headphones turn it into sound. Some crystal sets also have an audio signal filter or “integrator” capacitor. This capacitor, about 0.01 microfarad, is placed across the headphones. However, in my crystal set, it didn't do anything useful, so I left it out. Leaving out parts is good way to find out what they do.

Try leaving out the LC circuit and just connecting the diode and headphone to the antenna and ground. At my house all I could hear was faint static that sounded like power line noise. That implies that power lines generate the biggest AM signals over the entire radio spectrum. In any case, without the LC circuit, I heard no radio stations.

The Jamestown crystal diode

To make my crystal detector out of sulfide ore, I melted a puddle of solder about 3/8 inch wide on a piece of PC board. Then I used tweezers to press a bit of ore into the puddle so that, when it hardened, half of the crystal was exposed. Next I soldered a tiny loop of copper wire onto a pad on the board for a cat whisker. My first diodes used loops of fine copper wire as "cat whiskers." The copper wasn't springy enough to poke into the crystal with enough force for reliable performance. It worked well, but with a little vibration, it quickly died.



A safety pin pushes the copper whisker against the galena.

In my next diode I made a copper ring cut from the end of 1/4 inch copper tubing that served as a deep "tub" of molten solder into which I could push the galena. Bob, NØRN, told me that when he was a kid, he used safety pins as cat's whiskers. Sure enough, the spring-loaded safety pin produced plenty of force and solved the mechanical problem.

Carbon steel is a semiconductor

When I first put my diode with the safety pin cat whisker into a crystal set, it was stone silent - nothing. No matter how I moved the sharp steel pin around on the galena, the headphones were dead. The pin happened to strike the solder at the edge of the galena and the crystal set came to life with music from KBCU, our loudest local AM station. At first I was mystified. The steel pin rectified well against either solder or copper. The signal was perhaps only 2/3 as loud has it had been with the copper-to-galena diode, but it was much easier to adjust.

It turns out that steel is a carbon-iron semiconductor compound called cementite. The surface of hardened steel is a crystal, perhaps not radically different from the galena (lead-sulfide) crystal. Carbon has a valence of four, just like silicon or germanium. So, if you want to build a crystal set for your kids, ***you don't have to mine galena. Just use a safety pin pressing against copper or solder.***

Another surprise for me was that copper-to-copper, solder-to-solder, or solder-to-copper junctions also rectify and produce weak signals. The contact between the two metal surfaces must be extremely light - just barely touching. This phenomenon is poor for making crystal sets, but it's a warning about bad contacts in electronic equipment. ***Cold solder joints and loose screws can fill your circuit with accidental diodes.***

Copper cat whiskers work best

As shown in the diode construction diagram, I used acid core solder to attach a piece of copper wire onto the end of the pin. Now the contact point of my diode is between the semiconductor galena and copper rather than semiconductor steel-to-semiconductor sulfide ore. I connected my crystal set to the center conductor of my 40-meter dipole coax and my station ground. I scratched the copper whisker around on the sulfide crystal and suddenly I was again hearing our local station. Using commercial 8-ohm headphones, it was almost painfully loud. Too bad KBCU is mostly rap music.

Where is the P-N junction in these crude diodes?

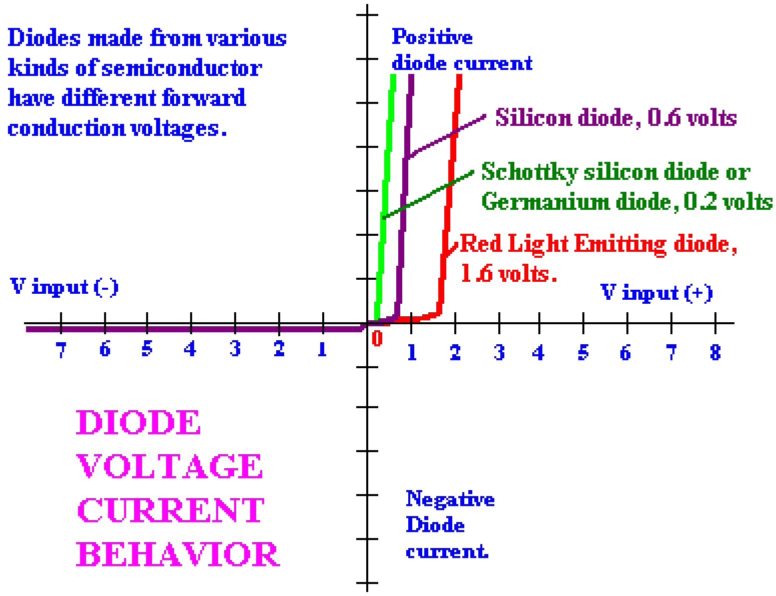
If you are a thoughtful person, you must be asking, "Where is the P-N junction with the impurities imbedded in the pure semiconductor and all that?" It turns out that you can make crude diodes by throwing together pretty inferior materials. For example, pure galena crystal consists of lead and sulfur that have valences of 2 and 6, that sort of average to 4. But there are also all those other atoms in typical galena ore. These impurities, like silver or copper, have valences like plus 1, while other transition metals like tin have valences of 2 or 4. Let's just assume that because of sulfide (valence 6) and arsenic (valence 5) salts, my ore is a N-type semiconductor. So where is the P-type semiconductor? It turns out that if you press a metal against N-type semiconductor, metal ions will migrate a few microns into the N-type crystal and make a tiny P-type region surrounding the contact point of the "cat whisker."

As you might expect, the disadvantage of such crude diodes is that P-N junction is quite fragile. That is, the check valve function only works with very low voltages and extremely small currents. The P-N junction is easily destroyed if you put large reverse voltage across it or try to pass large currents through it.

I tried measuring the volt/ ampere characteristic of a steel/ copper diode. As you would expect, it looks like a short circuit on an ohm-meter. I studied it carefully with a high impedance multimeter and 10 megohm series resistors, but it still looked like a short circuit. I believe I just learned that ***the world of RF detection is quite subtle***. At least the legend about POWs in WWII making radios out of barbed wire and razor blades is starting to make sense. The razor blade must have been the semiconductor.

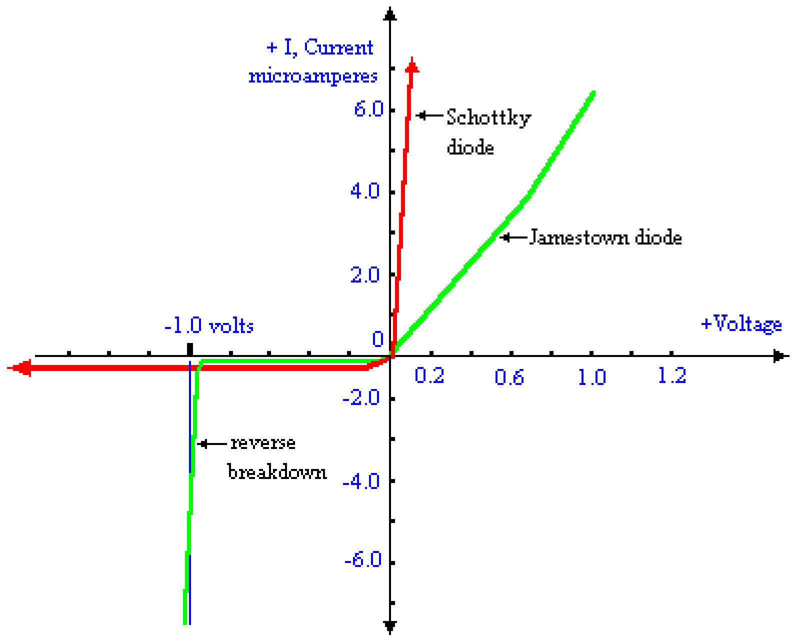
Commercial diodes

A perfect diode would have zero resistance in the forward direction and infinite resistance in the reverse direction. As you can see in the graph below, no real diode is perfect and all diode types have different forward voltage characteristics, depending on the semiconductors used to make them. The reverse characteristics of commercial diodes are all excellent and similar, but each type of semiconductor diode needs a certain level of forward voltage before it will begin to conduct. In the graph below relative current is plotted on the vertical axis. Commercial diodes handle hundreds of milliamperes or even hundreds of amperes. As you'll see below, homemade diodes can barely handle microamperes of current before they fail.



Volt – ampere characteristics of homemade diodes

Once I had my cat whisker adjusted, the Jamestown diode was just as loud as the 1950 crystal set diode. I tried substituting a modern, hot-carrier Schottky diode for the crystal. In theory, a Schottky should be comparable to a point-contact diode. Schottky diodes are commonly used as detectors in relatively modern equipment. I was surprised to discover that the modern Schottky produced as big a signal as the crystal diodes, but no better. If these diodes perform the same, how do their volt/ ampere characteristics compare?



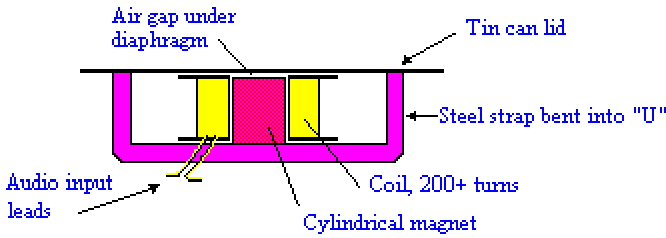
Volt / Ampere curves for the Jamestown diode and a commercial Schottky diode

At big voltages and currents, (milliamperes), the Schottky diode acted as you would expect: It passed big currents (milliamps) above 0.2 volts forward voltage and leaked only 100 nanoamperes with reverse voltage. Notice that if the Schottky were "perfect," the red line would be directly on top of the vertical axis upward, and directly on top of the horizontal axis to the left. However at very low currents, microamperes, the commercial Schottky was pretty nearly perfect with a transition right at zero volts.

In contrast, with big currents the Jamestown diode behaved like a resistor in both directions. At first it wasn't obvious to me that it could rectify anything. The curves above show the behavior of both diodes at tiny current levels, microamperes, using a 1 megohm (one million ohms) load. For tiny currents, the Schottky and Jamestown diodes were both strongly non-linear at the zero current, zero voltage point. The surprise for me was that, for reverse voltages, the Jamestown diode broke down abruptly at minus one volt. It's no wonder it conducted so well in both directions with a "low" resistance 10K ohm load. This abrupt, reverse breakdown is called ***avalanche breakdown.*** When it occurs with big currents it usually destroys the diode. As we'll see in chapter 8, some diodes called ***Zener diodes*** are designed to breakdown at specific voltages without being destroyed.

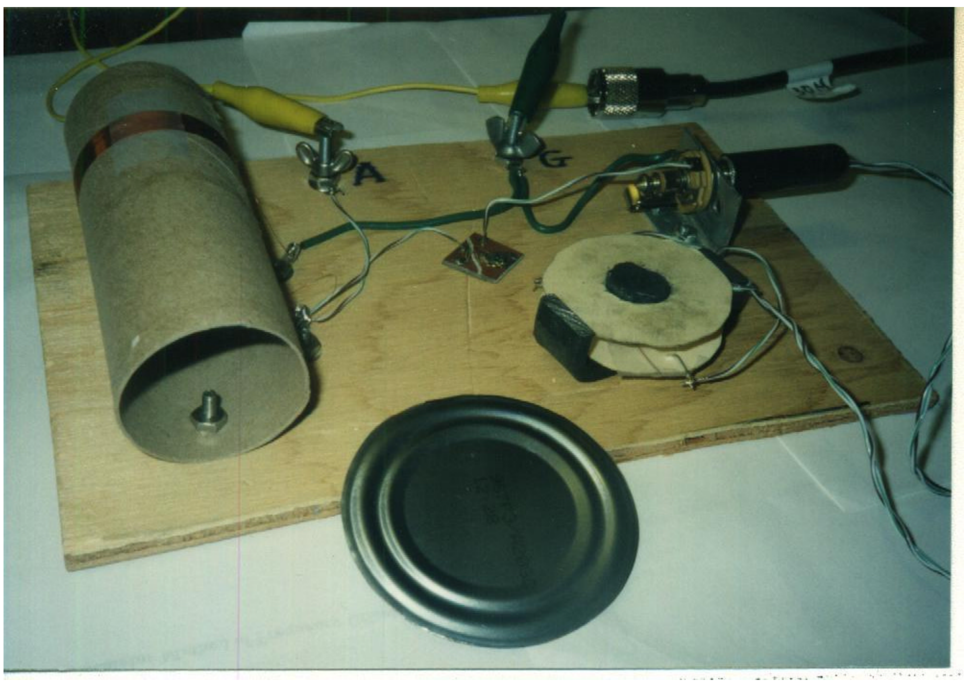
A homemade headphone

Building my own headphone was the hardest part of my crystal set. A headphone uses a high impedance coil of wire to make a magnetic field proportional to the audio signal. The changing field pushes and pulls against a thin steel diaphragm to produce sound vibrations. Even if you decide to build one of these headphones, I strongly suggest you buy a good pair of headphones so you will have them for your ham rig. Also, with commercial headphones the speech and music will be perfectly clear and loud, even with a safety pin diode.



The Caribou headphone

A cross section of my homemade headphone is diagramed above. Its construction is basically the same as old-fashioned high impedance headphones. Yes, the sound is tinny. What did you expect from a headphone diaphragm made from a tin can lid? The coil is hundreds of turns of #36 wire wound on a paper coil form. Inside the coil is a cylindrical magnet I took out of an old loudspeaker. A piece of steel strap conducts the magnetic flux around to the edges of the lid. The magnetic force holds the lid on. By completing the magnetic circuit, the magnetic force is concentrated in the gap between the tin-plated steel and the magnet.



Crystal set showing homemade headphone. The tin can lid diaphragm has been removed.

I started out using a small magnet from an old loudspeaker, but that felt like cheating. Would Heinrich Hertz have been able to use a loudspeaker magnet? Anyway, it seemed to me that the magnet wasn't essential. Why couldn't the coil just magnetize ordinary iron? I tried substituting a big steel nut of the same size. Sure enough, it worked, but the sound was too faint to be audible in the crystal set. However, when I plugged the homemade headphone with the steel nut into my shortwave radio, it was surprisingly loud. Not Hi-fi, mind you, but loud. No, for a sensitive headphone a magnet is needed to overcome the hysteresis.

Hysteresis

What's hysteresis, you ask? Whenever iron is magnetized with a DC coil, the tiny "magnetic domains" in the iron line up to make a big magnetic field. But when the DC current is shut off, some of the magnetic domains remain aligned and leave a residual field. To magnetize the iron in the opposite direction, a current of the opposite polarity must first overcome the residual field. This means that hysteresis interferes with the sensitivity to weak signals. Since crystal sets are powered by the radio waves themselves, sensitivity is vital. A magnet is needed to overcome the hysteresis and "bias" the magnetic field so that it always operates in one direction. I could magnetize iron with a DC powered coil, but then to be a purist, I would need to build a homemade battery. And I would need to smelt and extrude my own copper wires. (Forget I said that.)

I had a sudden inspiration. I dug around in my rock collection and found a piece of magnetite ore from a mine dump at Caribou, Colorado. Magnetite is a specific iron oxide, Fe_2O_4, that retains a magnetic field. I machined the magnetite with my bench grinder into a small cylindrical magnet. Unfortunately, the grinding and heat ruined the magnetism. However, fixing it to a big, heavy permanent magnet, I was able to put my magnetite in a strong magnetic field. Then I banged on it firmly against my anvil. Believe it or not, that abuse restored the magnetic field. Behold! - The completed Boulder County rock and toilet roll radio!

How does it perform? Well, frankly the homemade headphone is pathetic and needs lots of R&D. The sound is plenty loud when plugged into a real radio, but installed in the crystal set, I can just barely hear the rap music. Perhaps if I had a thinner steel diaphragm, a headphone for each ear, optimum impedance matching, better craftsmanship, and other refinements, it might approach a commercial headphone. In other words, for serious listening, buy a good headphone! And, after your done playing with homemade crystals, I suggest you buy some silicon diodes. Type 1N914 or 1N4148 diodes work great in this radio. ***They don't work better than the diode made from sulfide ore***, but they are smaller, more rugged, and don't need to be tweaked.

This crystal set article was first written in 2002. I revisited crystal sets in 2006 and discovered some surprises.

****

CRYSTAL SETS REVISITED

Or, electronics makes us humble

A friend of mine, Jack Ciaccia, WMØG, suggested to his grandson, Rutger Koch, that a broadcast band crystal set like the one described above would be a good science project. Rutger did a terrific job of duplicating the set and the homemade diode. The only obvious difference I could see was that he had used the 5-inch cardboard oatmeal box coil form that I had recommended. But when Rutger put it together, it was totally inert - no sound.

Rutger consulted his granddad who also couldn't find anything wrong. Jack brought the crystal set over to my house and we two old hams scratched our heads and still couldn't find the problem. After five minutes of swapping parts, I dug out my prototype crystal set from the closet and there was the answer: If you go back to the drawing of the homemade diode made out of the piece of PC board and a safety pin, you will see that there is a notch carved in the PC board between the anode and cathode. Rutger had forgotten to cut the copper sheet between the anode and cathode. We cut the copper sheet and it worked perfectly. ***Our eyes only see what they expect to see!*** Don't take the obvious stuff for granted. By the way, Rutger's science teacher refused to believe rocks could serve as diodes until he put on the headphones. This ancient technology is all new for whippersnappers.

Does a bigger diameter coil work better?

When I first built my AM broadcast band crystal set described earlier, a toilet paper roll was the largest diameter cardboard form I could find. Eventually, my wife and I finished up our own Quaker Oats and I had saved the box. Although I hadn't noticed any obvious superiority in Rutger's crystal set over mine, I decided to wind a big coil like his and see if my magnetic antenna theory had any validity. That is, the big coil might pick up the magnetic wave components and produce a bigger signal. Perhaps it would be directional like an AM band ferrite core antenna. Since my crystal set had previously been converted to 10 meters (see the next article, below), I just took off the loop antenna and left the 140 pF variable capacitor on the set. Then I installed the oatmeal box coil with 20 turns on it. Since I already knew that ordinary silicon diodes have the same sensitivity as the rock diodes, I used a 1N4148 diode and my large, low impedance headphone.

Oatmeal box crystal set - operating without a ground or counterpoise

I hooked up the 40-meter dipole antenna, "A" above, and used the 30-meter dipole as a counterpoise connected to the ground post, "G" above, just as I had before. This was the same way Jack and I had hooked up his grandson's set. The silence was deafening. Good grief! Now what?

After some flailing about, I accidentally disconnected the counterpoise wire as shown above, just leaving the antenna wire connected. Voila! It not only worked well, the tuning was precise enough to tune in my local NPR station at 1490 KHz with excellent selectivity and signal strength. By "excellent" I mean, it was not only loud and clear, I could only hear the one station. Precise tuning with a crystal set! How is this possible? When I padded the relatively small 140 pF variable capacitor with 100 pF fixed capacitors, I was able to tune the down the AM broadcast band and uniquely select other stations. The farther down the band I went, the less precise the tuning. I believe the loss of precision at the low end of the band is related to the relative signal strength of the stations. Also, I obviously needed more turns on my coil. If I were starting over, I would use a larger variable capacitor, e.g., 365 pF.

How does this work?

The selectively tuned crystal set seems to be an example of a circuit that doesn't seem to have a complete circuit loop. Without a ground or a counterpoise, where does the current go? I assume that there is some form of capacitive coupling that completes the loop, but I must say this coupling isn't obvious. As you can see, RF sometimes makes current loops that are hard to identify.

Using the backyard antenna as a counterpoise didn't work with the big coil, but I found that clipping the ground terminal, "G" to my station ground did improve signal strength. However, it ruined the precision of the tuning. Hooking the station ground and the counterpoise antenna together still worked, but was weaker and produced poor tuning. Using the backyard 30-meter antenna as the sole antenna, worked just as well as the 40-meter antenna.

I probed with my oscilloscope at various places on the circuit to get some answers. As soon as I attached the oscilloscope probe ground wire, it behaved just like the grounded hook up, just as you would expect. The signal strengths across the coil were on the order of 0.40 volts peak. The power delivered to the 8-ohm headphones was about 20 milliwatts peak. The most surprising observation was that, when I just left the scope probe and its ground wire lying six inches away from the crystal set, I could see the modulation on the screen tune in and out as I selected specific stations. The voltage signal was only slightly less than when it was directly connected. Apparently there is a large electric field surrounding the crystal set.

As for the magnetic antenna effect, when tuned into a specific station, there was no hint of directionality as I rotated the coil from north to south. In conclusion, the advantage of the big coil seems to be a dramatically higher Q (higher quality resonance). But to observe this advantage, the headphone must be powered exclusively by the signal ringing across the coil. If the current is passing from antenna to ground, the L-C circuit has much less effect on the tuning. And finally, (apparently) when the ends of a high Q coil are connected to two equal antennas, the signals cancel each other.

Perhaps you have your own theories about these phenomena. Feel free to pontificate. Obviously we old hams don't know everything, even about the simple stuff! Keep your brain engaged and question authority.

****

RECREATING HERTZ'S RADIO EQUIPMENT

Most of what I've read about the history of radio was written by non-engineers. They describe the revolutionary apparatus invented by our heroes using old-time radio terms like "earth resonances," "aether," and "coherers." They tell us how far it transmitted, but they give us only the faintest clues about how the gizmo actually worked. Was it a spark gap transmitter? A high-speed alternator? What the heck was a "Tesla oscillator" anyway?

In 1884 James Maxwell published four equations that quantified and connected magnetism with electric phenomena. These equations also predicted the existence of radio waves. The changing magnetic and electric fields related to each other with sine functions. So, once physicists had the equations to stare at, it wasn't too huge a leap to conclude that sinewave-shaped electric and magnetic fields would generate each other in an oscillation and radio energy would propagate through space.

In 1889 Heinrich Hertz, a physics professor at the University of Bonn, Germany, was the first to demonstrate radio waves in the laboratory. Of course he might have done this in 1884, or 1887, depending on which website you visit - ah, the glorious information age!

That's fascinating, but **HOW** did he demonstrate radio waves? Using 1880 technology, that could not have been easy. How did he know he was detecting waves and not just magnetic coupling from one coil to another? Or if his "antenna" was capacitive, how did he know he wasn't observing capacitive coupling? ***If I were skeptical about the existence of radio waves, but I understood the full implications of Maxwell's equations, I would be convinced if I could see communication across a distance greater than one wavelength.*** A minimum of one wavelength means that "the alleged electromagnetic wave" would change from magnetic to electric field energy then back again at least once. ***Of course, I would also want to see evidence of standing waves and a way to measure frequency.***

Demonstrating Hertzian waves

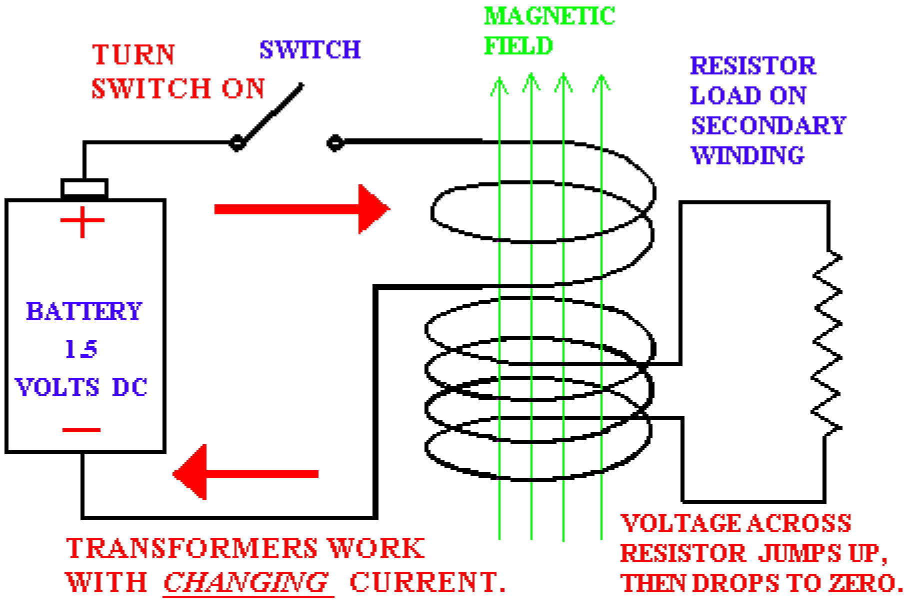
Suppose you were living in the year 1884 and Maxwell had just predicted the existence of radio waves. Using components available in your time, how would you generate Hertzian waves and get those waves named after you instead of Hertz? If you are able to generate radio waves, how could you prove to a skeptic that you had actually done it? Hertz managed this feat and apparently his demonstration was convincing. Otherwise the unit of measurement for frequency would not be the ***Hertz***. One Hertz (Hz) equals one cycle (one complete oscillation) per second.

When I first had the idea of playing at being Dr. Hertz, I was not able to find a description of his apparatus. That was a good thing because it forced me to invent my own method to demonstrate Hertzian waves. If you already know enough about electricity to be able to handle the challenge, then get out your 1880 hardware and build a transmitter and receiver that will transmit at least one wavelength. If you don't know how to begin, keep reading.

Transmitting and receiving as simply as possible

The only detail of Hertz's apparatus I found described was that he detected his waves by means of a loop of wire. The wire had such a large current and voltage induced into it by the radio waves that a visible spark jumped across a gap in the circle of wire. Wow! It must have been a big radio signal that would induce that much energy into a loop of wire. And if the signal was that big, how far away from the transmitter had the loop been? I suspect the signal strength had to be big and loop had to be very close, like a foot or two away.

If were a skeptic who already knew about Faraday's transformers, this demonstration would not convince me. How would I know that radio waves had propagated across the one foot distance to the loop? Maybe all I was seeing was a big magnetic field that reached from one coil to another.



Maybe Hertz's demonstrator was just a transformer?

A transformer is a *magnetic* device that works by transmitting a changing magnetic field from one coil to another. Coils (inductors), convert the energy of an electric current moving through a wire into magnetic field energy that hovers in a cloud like region around the coil. If a second coil is close enough to the first coil to be inside the magnetic "cloud," then if the magnetic field is changed, then an electric force, a voltage, will be generated in the second coil. Transformers and the above diagram were discussed in chapter 2, page 26.

Inductors store magnetic energy in the space around them so long as current is passing through the coil. The energy will remain in space so long as the current keeps flowing in the same direction through the coil. But when the current stops flowing, the magnetic energy becomes "stranded" in space. The magnetic energy then returns to the coil and induces a voltage in that coil in a direction that would force the current to continue flowing. That is, the induced voltage will try to keep the status of the magnetic field and coil the same. If the current is gone for good, then the magnetic field will collapse completely and the energy will dissipate into the coil and any circuit connected to it. But if the first coil is open circuit, and even high voltages cannot restore current flow, then the field will collapse into the second coil. If the induced voltage is able, it will cause current to flow in the second coil to maintain the field. Or, as in the case of Hertz's loop detector, the induced voltage caused a big spark to jump across a gap where the resistor is located in the illustration above.

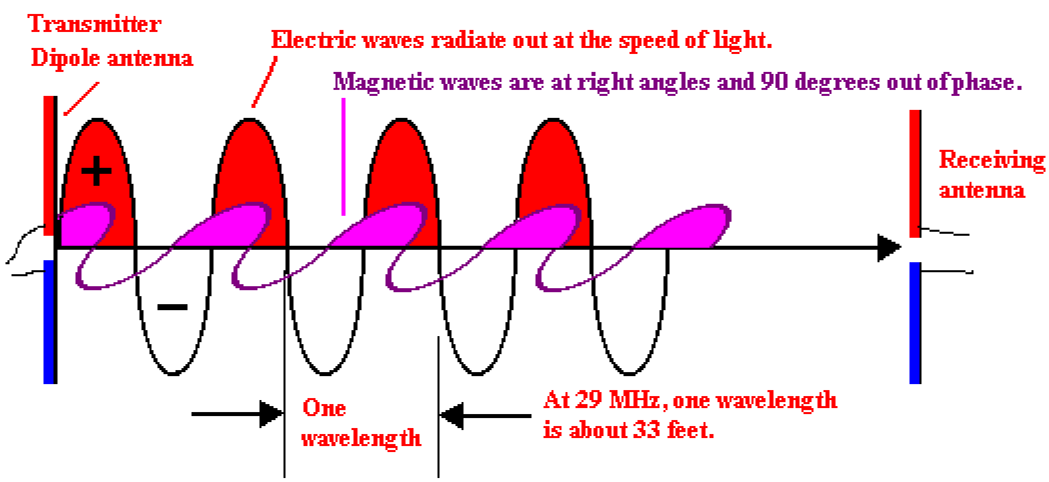
But I thought transformers were always made from iron

You may be thinking that transformers are not air-filled flimsy coils but rather large iron things like the big steel cylinder on the power pole in the alley behind your house. Yes, those big iron things are transformers but they are designed for low frequency power lines. An AM radio frequency is on the order of a million Hz frequency, while the power company supplies current at 60 Hz. Therefore, instead of having air between the two coils, power transformers have iron. The magnetic field from the coil magnetizes the iron temporarily. Recruiting iron and turning it into a magnet increases the magnetic field a thousand times or more. With a huge magnetic field stored in the iron, the power transformer can transmit big amounts of energy with only 60 direction changes per second. A similar transformer without an iron core could transmit the same amount of energy, but would have to repeat the magnetic field cycle perhaps a thousand times more often to transfer the same amount of total energy.

Why do you suppose the power company doesn't use 1million Hertz and do away with all that iron? After all, at one million Hertz the RF voltage would still be a burn hazard, but it could not electrocute anyone and would be considerably safer. Unfortunately, at one million Hz the power lines would act like antennas and radiate the energy into the sky instead of delivering it to your house. This is the same reason hams object to using the power lines to distribute high-speed Internet connections (BPL). BPL broadcasts noise all over the shortwave bands and makes them nearly unusable.

How far should it transmit to demonstrate the existence of Hertzian waves?

To be sure that the waves are Hertzian and not just magnetic fields, I would be impressed by the demonstration if the detector (the receiver) were more than one wavelength away. A wavelength is the distance that a radio wave travels during the time it cycles from magnetic field, to electric field, and back to magnetic field.



The speed of light is 186,000 miles an hour, or 300,000,000 meters per second. (A meter is about 39 inches) A wavelength is the distance a wave travels while going through one cycle of magnetic to electric energy conversion. The wavelength of the forty-meter amateur radio band (7 MHz) is obviously 40 meters. It turns out that typical, (total) antenna lengths for radio transmitters are either one half wavelenth or one quarter wavelength. On 40 meters, a typical vertical pole antenna is one quarter wavelength or 10 meters (33 feet) tall. The 10 meter ham band extends from 28.0 MHz to 29.7 MHz.

What exactly is the frequency in Hertz of the 10 meter ham band? To convert wavelength to frequency, divide meters per second of light speed by the wavelength:

$$f = c / \lambda$$

Where f represents frequency, c represents the speed of light and λ represents wavelength

Speed of light/ wavelength = Frequency in Hz.

300,000,000 meters/ second / Ten meters = 30 Million Hz (30 MHz) frequency

Remember that the AM radio band extends from 550,000 Hz to 1.7 MHz. Channel 2 television starts at 54 MHz . So the 10 meter ham band is roughly halfway between AM Radio and TV.

Getting back to the Hertzian demonstration, if I wish to transmit one wavelength, on 40 meters, my loop would have to be 132 feet away from my transmitter. Frankly, I don't think Hertz's loop detector will work at that range. And if it did, I would be arrested for using a transmitter that powerful. As a rule of thumb, the American FCC will not object to experiments like this if the radio waves don't go past 50 feet at easily detected signal strengths. On the other hand, maybe I could use a higher ham band like 10 meters. Now I only have to go 33 feet. This is better, but the loop is still more than one foot away. If I go up to UHF frequencies, the wavelength could a foot or less, but those frequencies would be hard to generate and harder to measure with 1884 technology. I have since been told that Hertz actually used 4 meters wavelength for his demonstrations.

Designing the 10 meter transmitter

For the sake of symmetry I used two identical LC circuits for my transmitter and receiver. To imitate what little I knew about Hertz's apparatus, I used one loop of wire about a foot in diameter. I knew from experience that seeing sparks on the receiver side was hopeless, so I also made the receiver an LC circuit that I knew would trap an oscillation from the transmitter.

At this point, you may want to depart from the year 1880. It all depends on how pure you wish to be playing the historical game. For a capacitor I used a modern 140 pF variable capacitor. That way, I could adjust the capacitance and tune the oscillation to a particular frequency. If I wanted to be a purist, it would not be hard to make a homemade capacitor made from sheets of metal with paper for insulation between the plates. Personally, I was confident that such a capacitor would work. I just didn't want to spend hours to make one.

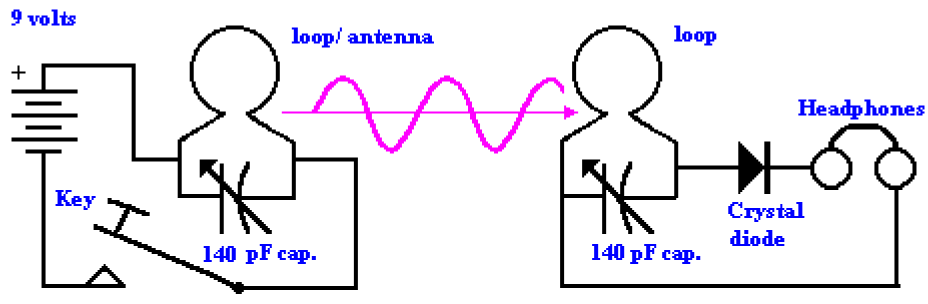
My first problem was how to start the oscillation in the transmitter LC. In theory, by shorting a battery across the loop, it will charge the loop with a big current limited only by the internal resistance of the battery. Then when the battery is removed, the coil's magnetic field will discharge forcing a voltage to appear across the capacitor. The loop will then be shorting out the capacitor and the oscillation will begin.

Like the variable capacitor, this project will go faster if you use modern tools to make sure your components are working. For example, to see if my transmitter was really transmitting, I used a ham band receiver tuned to ten meters. Sure enough, when I clicked the battery on the capacitor terminals, I could hear a click in the receiver loudspeaker. And when I tuned the capacitor, I could get the sound to reach a sharp maximum volume at a specific setting of the capacitor. Of course, if this were 1880, I would have to do everything by guess, trial, and error. Those old guys were darn sharp.

Designing the receiver

The next problem was how to detect when the receiver loop was oscillating due to waves from the transmitter. What to use for a detector? The 1880 solution would be to use a detector called a ***coherer.*** Coherer detectors were developed for the wire telegraph. By the time a signal had propagated miles down a telegraph wire, signal was often too weak to close a mechanical relay. Coherers were used to "amplify" a weak Morse code signal. They were described in chapter 1, page 10. Unfortunately, coherers are a low frequency device, under 20 Hz. They are suitable for detecting a weak DC Morse code signal on a cross-country telegraph wire, but will not respond to the modulation in the audible range. I doubted that they would be useful for tiny radio frequency signals. Being lazy, I didn't build one to find out. Besides, my crystal detector made out of local rocks certainly fit the 1880 criterion. I decided to build a 10 meter crystal set.

Are headphones 1880 technology? Yes, just barely. Alexander Bell built his first telephone in 1879. He used a headphone designed like the homemade device described earlier. Actually, for my 10-meter receiver I used old commercial high impedance headphones instead of the homemade earphone. I couldn't afford to waste any sensitivity.

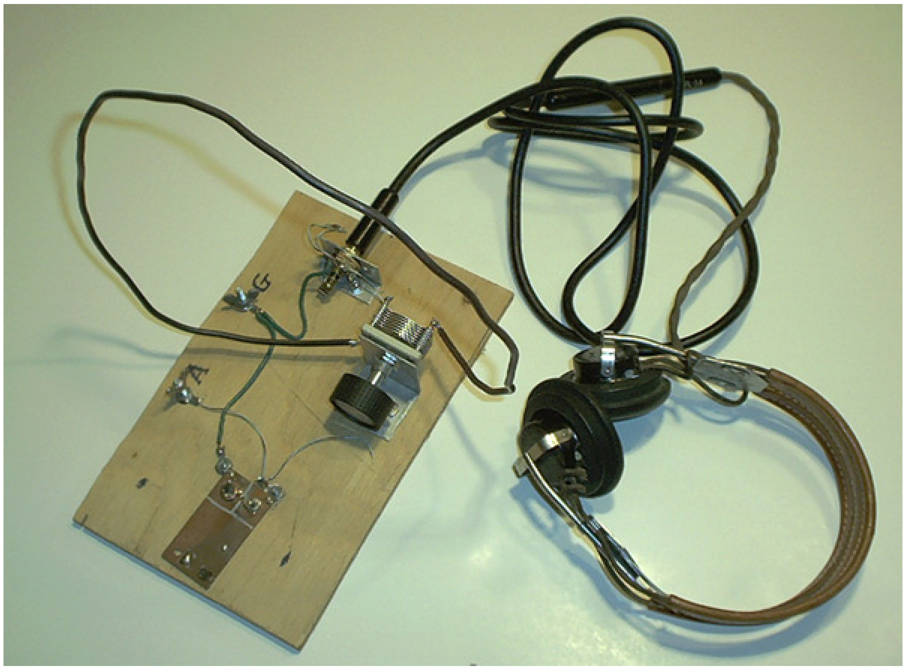


1880 Transmitter

Crystal set receiver

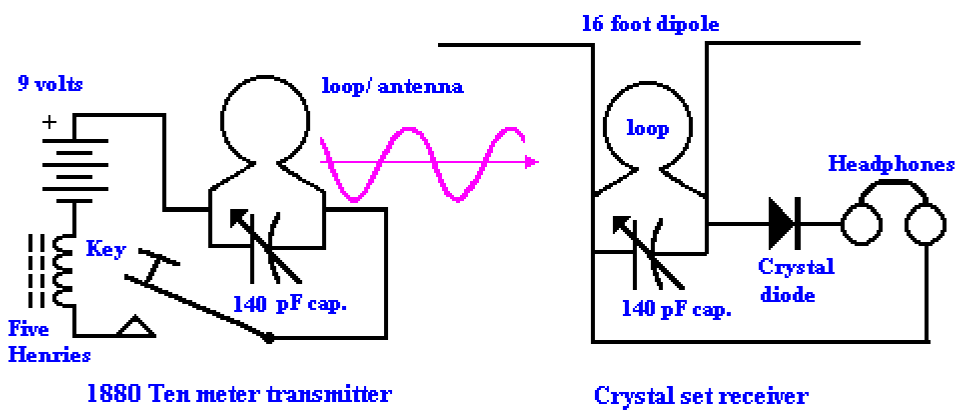
The simplest 10-meter communicator

I began with identical loop antenna/ inductors about 1 foot diameter for both receiver and transmitter. Instead of Hertz's spark gap, I put 140 pF variable capacitors across both loops to give me tunable LC circuits.



10-meter crystal set receiver

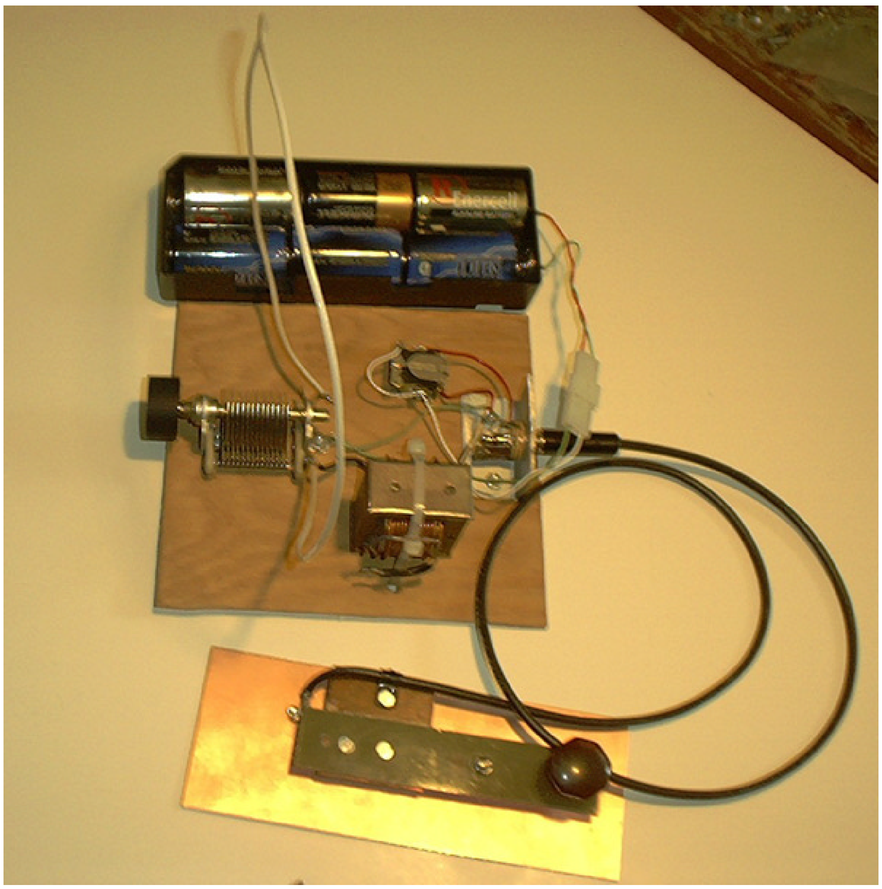
To charge the LC circuit, I used a crude ***spark gap***. I just touched battery terminals across the LC circuit while listening to the crystal set. I positioned the crystal set a foot away from the transmitter loop. It wouldn't be "radio" communication, but it would at least tell me if I was on the right track. I made sparks on the transmitter loop while tuning the capacitor. When tuned to just the right spot, I could suddenly hear obvious clicking in the headphones. I was surprised how sharp the tuning had to be. The big loop had relatively low inductance, so the capacitor had a tuning range of over 30 MHz. Tuning was probably not sharp by modern standards, but the adjustment was critical. In any case I achieved a range of 12 inches from the transmitting loop. Progress! Well, it's much farther than the obvious range of a refrigerator magnet.



Improved spark gap transmitter & receiver

By adding a big iron core inductor in series with the battery, I got a much bigger, more sustained spark and a much louder signal in the crystal set. The inductor was the primary of an iron core filament transformer that I had in my junk box. The secondary of the transformer was left open circuit. Actually, I tried several transformer windings and inductors until I found one that gave me the biggest visible spark. The battery was six D-cell alkaline batteries in a plastic battery holder from Radio Shack. My telegraph key was made out of two pieces of printed circuit board separated from each other by a piece of wood. (See Chapter 9.)

Using an extra inductor in this way is analogous to the automobile ignition system described in chapter 2. Without the big inductor, the only inductance to store energy was the single loop of wire. So when I tapped the battery wire on the LC circuit, very little energy was stored in the capacitor and inductance. The spark in a spark gap transmitter happens when the wire charging the inductor is broken, just like opening the breaker points in a car ignition. The more energy stored in the system, the bigger the spark when the connection is opened.



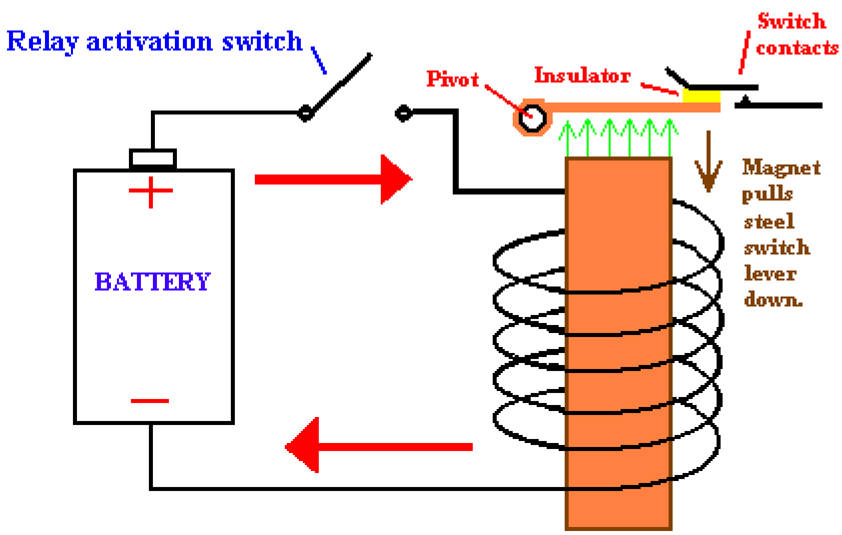
10 meter sparkgap transmitter

I drafted my XYL to listen to the headphones while I moved the transmitter across the room. (Wives are known as “XYL” in Morse code. XYL stands for “former young lady.") Now that I had the inductor and a larger spark, I got clear out to a range of 10 feet! I explained to Katie that she was doing the same job as Marconi’s assistant, Mignani. When Mignani heard the repeated “S” in Morse, he fired a rifle into the air. “Don't I get a rifle?” Katie asked.

A relay for automatic keying

Unfortunately, a wavelength at 29 MHz is 33 feet and I was still short. Hmmmm ... How to get the last 23 feet of range? First I added a relay to key the inductor. Yes, they had relays in

1880. The relay had nothing to do with extending the range, but it did enable me to continue doing experiments without having a Mignani to do the listening for me. You know, fire the rifle and all that.



RELAY CONSTRUCTION

A relay is an inductor/ electromagnet wound around an iron core. A hinged piece of steel is suspended by a spring near the iron core. When current passes through the electromagnet, the steel hinge is attracted to the iron with an audible "clunk." The hinge, in turn, mechanically closes a switch that can be entirely independent of the electromagnet circuit. In this way, one current can control an independent circuit. In my transmitter the switch contacts on the relay became my "spark gap." Every time the switch opened, a big spark jumped across the switch contacts, not unlike the spark in an automobile ignition spark plug.

Using a relay, I could use my electronic telegraph key, a "bug" set on "dots," to key the transmitter automatically. This homemade bug is described in Chapter 9. If you don't have one of those, you'll have to have your "Mignani" key the transmitter for you. With the transmitter making a continuous buzzing signal, I could move the receiver around my house. The signal sounded just like the automobile ignition noise that you sometimes hear in your AM radio.

More range = bigger, higher antennas plus bigger batteries

I could easily increase the range of the transmitter by using more and bigger batteries and a larger series inductor. If I really wanted to extend the range, I could add a dipole designed for 10 meters and put it up in the air about 50 feet. In fact, this is exactly what the early guys did – they made bigger and bigger transmitters and antennas. However this was 2002, not 1880. The trouble with using a wavelength of 10 meters is that, if I were to increase the effectiveness of my transmitter, I might easily hear it with my crystal set 33 feet away. Unfortunately, someone else might also hear it in Australia. That would be bad since sparkgap transmitters have been banned since 1927.

The simplest improvement I could make to the receiver was to add a 16-foot dipole antenna. The dipole consisted simply of two eight-foot wires soldered onto the sides of the receiver tuning capacitor. The dipole was oriented at right angles to the direct path to the transmitter. That did it. Now I could plainly hear the signal from the basement to the other end of the 2nd floor of my house, over 50 feet. That was well beyond one wavelength range. Success!

Looking for standing waves

To measure wavelength, I set up a long wire transmitting "antenna" about 50 feet long across the floor and upstairs. I reduced the transmitter batteries from 9 volts down to 3 volts. Then I turned on the transmitter. I took the dipole off the receiver and then used the receiver loop as a "probe." Walking along the wire, I was able to hear peaks and dips in reception every 6 feet or so along the wire. What I was hearing was "standing waves." When the RF current reaches the end of an open wire, it bounces back along the wire. The returning waves cancel and reinforce the outgoing waves making the peaks and nulls I was hearing. A large number of peaks means the wire length is different than one wavelength and the standing waves are complicated. If the wire was exactly one wavelength, I would hear just two peaks – they would be the two humps of a single sinewave.

Next I cheated. Since I already knew the frequency was 29 MHz, I calculated what the wire length should be for one wavelength. I trimmed the wire to exactly that distance and tried again. As expected, there was a single pronounced dip in the center of the wire. The sinewave signal was reflecting back and forth from one end of the wire to the other, with a dip, the zero crossing, in the middle. When the reflections don't come out even, you get many multiple dips and peaks.

Of course, knowing the answer before you start is not what Hertz experienced. He had to figure out all the details the hard way. Also, knowing the answer ahead of time biases the result. The exact alignment and distance of the receiver loop with respect to the wire were critical, so there's a chance I was just hearing what I wanted to hear. Craftsmanship and scrupulous honesty are essential when doing science. My frequency measurement obviously more needs work.

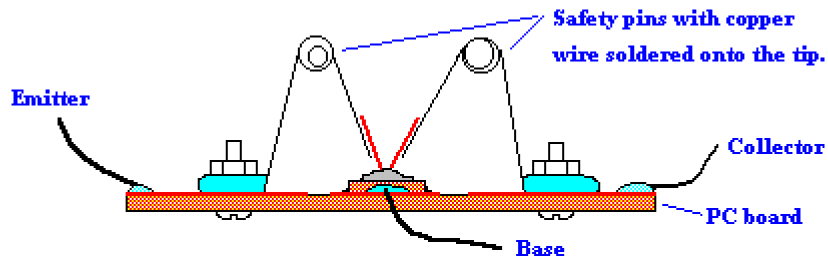
****

HOMEBREW TRANSISTORS

Here is another project you may enjoy. It will also introduce you to the basic principles of bipolar transistors. Schockley and Bardeen first invented bipolar transistors while working for the Bell Laboratories in 1947. Actually, I've read that the basic principles for field effect transistors were described in German patents from the 1930s. However, field effect transistors (FETs) were not developed into useful components until the 1970s. We shall first use an FET in chapter 6, so FETs are discussed there.

The purpose of transistors is to "amplify" small signals. Saying it another way, transistors control big currents or voltages using tiny control signals. Transistors can amplify a tiny signal that is audible only with sensitive headphones. They can make it strong enough to run a loudspeaker or even deafen a stadium during a rock concert. Alternatively, a transistor can use a little control signal to turn on a huge current and voltage. For example, an engineer at a power plant might push a keypad on a computer with a fraction of a milliampere of current flowing through the switch. This action is amplified and results in megawatts of power at hundreds of thousands of volts flowing toward a city.

The homebrew transistor

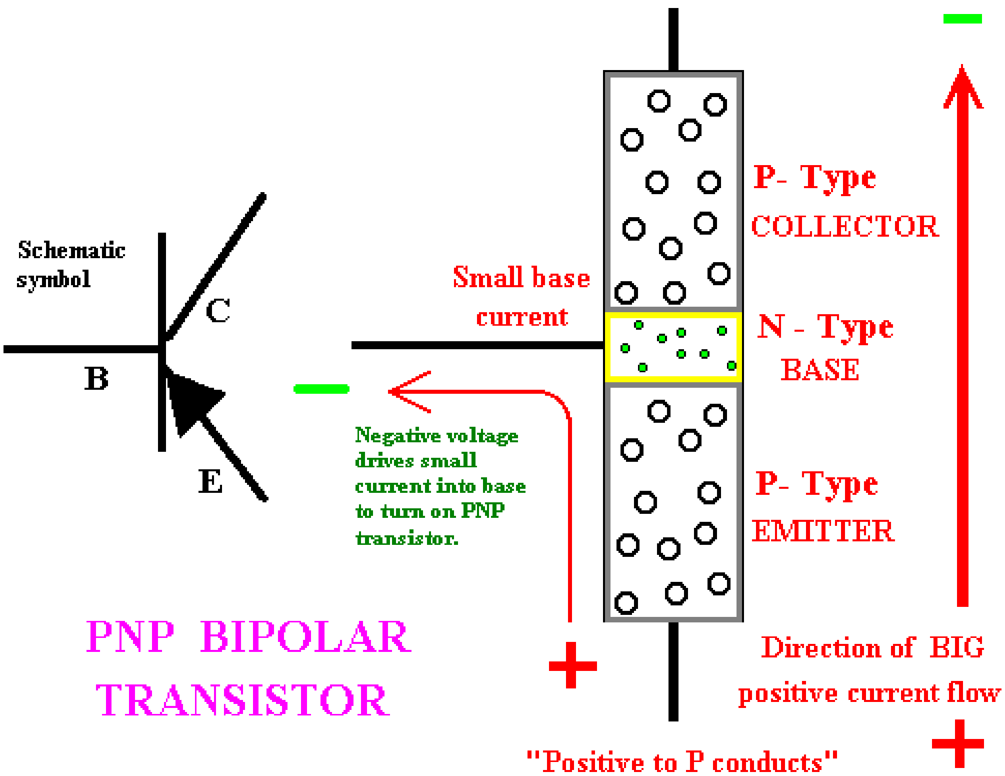


A point contact transistor

The first bipolar transistors were the "point contact" type. They were much like the galena diode described earlier. After my diodes worked so well, I wondered if I could make a transistor.

Idealized construction of a PNP bipolar transistor

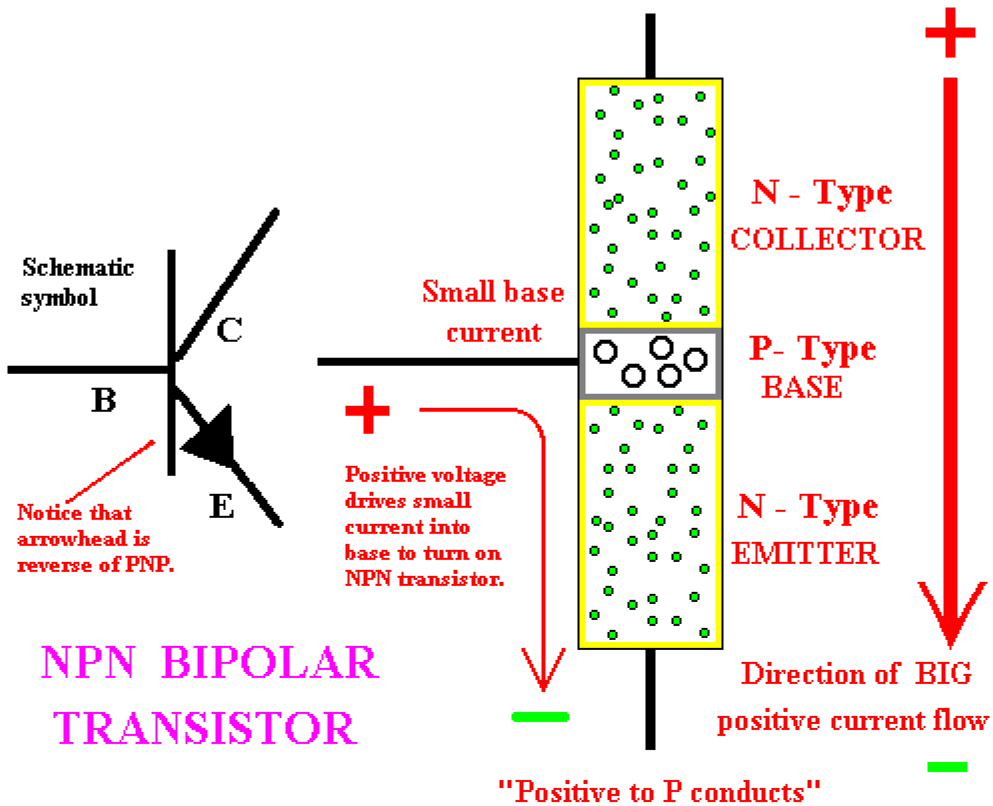
A ***bipolar*** junction transistor consists of two diodes made on the same semiconductor crystal.   In the experiment above, I am attempting to use galena as the N-type semiconductor. The diodes are connected back-to-back so that it will look like an open circuit from the terminals called "emitter" and "collector." The two copper points are supposed to touch the galena so close together that the tiny semiconductor region between the two points can be biased by the base current. The bias current is supposed to electrically convert the semiconductor region into a "conductor" and thus turn the two back biased diodes "ON." Don't forget to cut the copper sheet between the three terminals.



In theory, the metal from the copper points diffuses into the surface of the crystal and makes a tiny region of P-type semiconductor where the copper touches the semiconductor. Unfortunately the "emitter" and "collector" are identical and aren't optimized for their different roles as they are in commercial bipolar transistors. I also tried making a copper/steel/copper transistor, but without any measurable breakdown voltage, it just acted like a short circuit.

NPN transistors

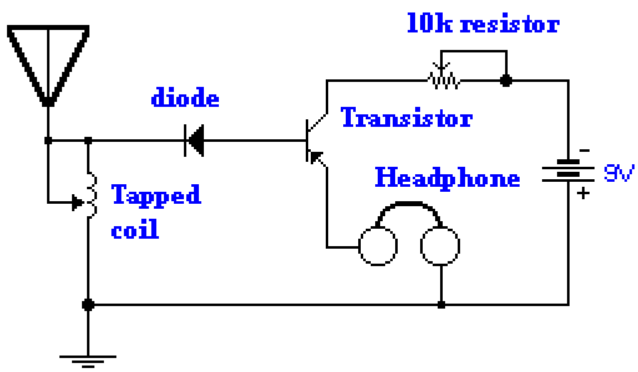
A nifty advantage of bipolar transistors is that they can be built two ways. By reversing the P-type and N-type semiconductors, an NPN transistor can be built that operates exactly like a PNP transistor, except all the polarities and current directions are reversed. The advantage of having two polarities is that the circuits can often be simplified by using both kinds in the same circuit. In practice, NPN transistors are usually slightly more robust and less likely to fail at high power loads. For this reason, the power amplifier stages in modern transmitters are almost always N-channel devices. On the other hand making an NPN transistor out of crude crystals and safety pins is inherently difficult!



Proving the homebuilt PNP transistor has gain

I measured the static volt/ ampere characteristics of my galena, point-contact transistor but couldn't show any gain with static DC currents even in the microampere range. Before I gave up, I thought I would try it as an amplifier in the crystal set. Maybe I could demonstrate gain in the subtle world of RF detection.

I decided my "transistor" was most likely to work as an ***emitter follower.*** In an emitter follower, there is no voltage gain, only current amplification. The load, the headphones, would be located between the emitter and the positive side of the battery. That would match the impedance between the high impedance detector and my low impedance (8 ohm) commercial headphones. Because galena diodes break down with typically 1 volt of reverse bias, I used a 9 volt battery with a 10K pot in series so I could limit the voltage on the collector to less than 1 volt.



Rock radio schematic

As I moved the emitter pin around on the crystal, a loud radio station abruptly appeared in the headphones. I disconnected the battery. Sure enough, the music was far louder with the battery connected.

An accidental crystal microphone

With the battery in place, but the signal diode disconnected, I heard a faint "sea shell" roaring sound - you know, like a live microphone. Tapping on the transistor assembly, I heard the scratching sound greatly amplified in the headphones. I seemed to have built a "crystal microphone"! I replaced the homebrew transistor with a real PNP transistor, a 2N3906, which amplified just as well but had no microphone-like characteristic.

Commercial crystal microphones consist of a lump of Rochelle salt held between two electrodes. When exposed to sound, mechanical vibrations cause the salt to generate a tiny audio frequency electric voltage. Crystals like this are called piezoelectric crystals and perhaps I had just built one. Piezoelectric quartz crystals are used to regulate frequency and are discussed in chapter 6.

Repairable transistors

While listening to a radio station, I slowly increased the battery voltage by lowering the resistance of the 10K pot. As the collector-to-emitter DC voltage rose higher, the volume increased higher and higher. I monitored the average collector to emitter DC voltage with a high impedance voltmeter. Then suddenly the voltage and sound crashed. I lowered the voltage again, but the sound didn't return. Good grief! I blew my transistor! No sweat. I just scratched the collector pin around on the crystal until I found a new "sweet spot" and I was back in business. Repairable transistors! Now there's a concept. After several trials I found sweet spots as high as 5 volts before the transistor died.

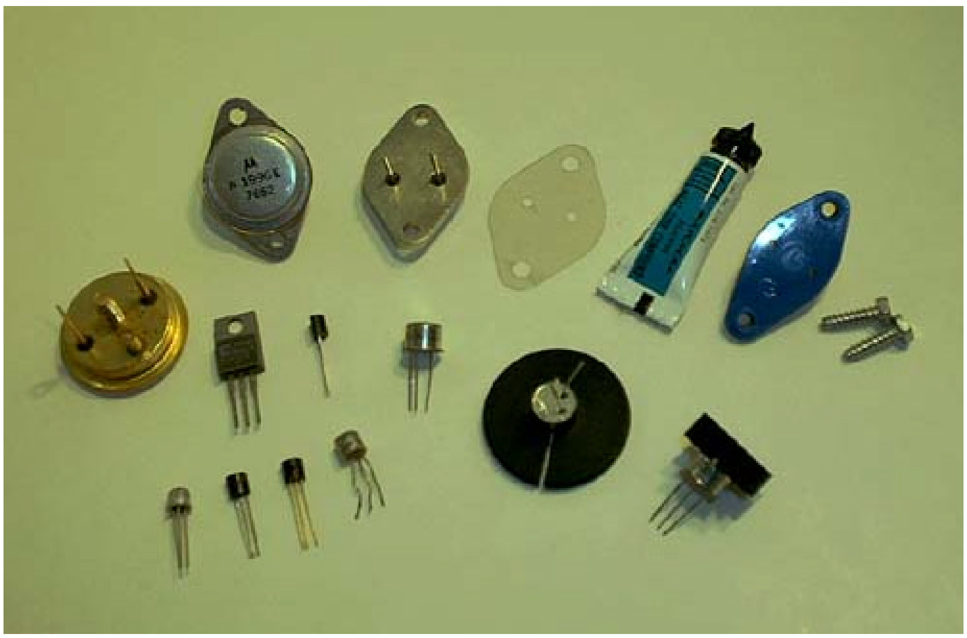
Now that I had an amplifier on my crystal set, I replaced the commercial headphone with the homebuilt "Caribou headphone" which I described earlier. You may remember that this headphone was made from a piece of magnetite ore and a tin can lid for a diaphragm. Sure enough, the sound was loud enough to understand actual words, rather than just distant music. More progress!

Is this gain or what?

In fairness, most of the "gain" or amplification I observed was just DC bias to the headphone which helps overcome the hysteresis of the steel components in the headphone. I replaced the transistor with a variable resistor so that the battery's sole function was to bias the headphone. That produced extra sound, especially for the homemade headphone that has such a weak magnet. After switching back and forth between resistor and transistor, the transistor was clearly louder, about 5 dB (decibels) louder measured with my multi-meter.

As I fiddled with the contact points, I eventually ruined the crystal. Under the microscope I could see copper smeared on the surface of the galena. When I built replacements, they never worked as well. I spent another morning trying to conclusively demonstrate gain by building an RF oscillator. I never got so much as a peep out of it, although as you would expect, a real 2N3906 PNP transistor worked just fine.

Real transistors



Above is an assortment of bipolar transistors. The little ones on the lower left include 2N3906 and 2N3904 devices used for low power oscillators and amplifiers. The medium sized transistors are used for amplifiers in the range of ¼ to 5 watts. Notice that two of them are wearing black aluminum heat sinks. The large power transistors in the back row are designed for power supplies and high power amplifiers. A mounting kit for case style TO-204 (formerly TO-3) transistors at the upper right consists of a mica insulator, silicon grease, and a socket with screws. The beautiful gold device is an obsolete high power germanium transistor.

In conclusion

Yes, Virginia, there are homemade transistors. But science that isn't reproducible isn't science. Without better basement technology, my homemade transistors have no future except maybe as microphones. Oh, well. The reward is the journey, rather than the destination. Keep thinking and dreaming!

****