Receivers: From Spark to 16-Qam

Course No: E03-023
Credit: 3 PDH

Edward Gellender, P.E.

CED CANADA
engineering.ca
CPD FOR THE PROFESSIONAL

Continuing Education and Development, Inc.

P: (877) 322-5800
info@cedengineering.ca

1.0 The Basic Foundations of a Radio

1.1 The Electromagnetic Spectrum

The electromagnetic spectrum shows the relationship between all electromagnetic waves by frequency. There is no lower limit, as the spectrum truly goes down to frequencies where no useful attributes have ever been discovered. Similarly, there is no upper limit. For this course, we are limiting ourselves to the part of the electromagnetic spectrum that is called the radio spectrum, roughly 10^3 to 10^{12} Hz.

The development of radio has slowly moved from left to right across the spectrum shown in Figure 1.

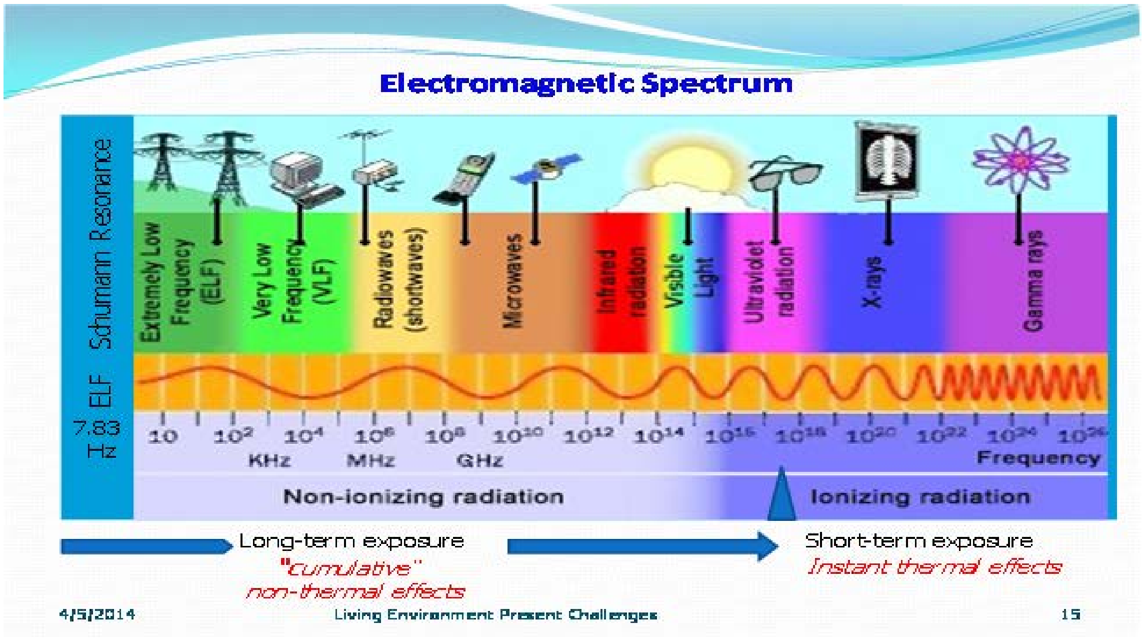

Figure 1 - A simplified representation of the electromagnetic spectrum

1.2 James Clerk Maxwell and the Aether

When experimenters learned how to create a vacuum, they quickly discovered that both sound and light waves propagate through the air, but only light waves will travel through a vacuum. They were concerned enough about why sound needs a propagation medium, yet light doesn't, to propose a theory that a yet-unidentified medium, named the Aether (pronounced like the organic chemical, ether), provides a medium for light waves to propagate through a vacuum.

The English mathematician James Clerk Maxwell was working with those experimenters, applying rigorous mathematical treatment to their findings. In 1861 he published what is now known as Maxwell's equations, showing no need for a propagation medium, at one time rendering the Aether a historic footnote, while creating a new understanding of the electromagnetic spectrum.

To this day, radio waves are still said to propagate through the Aether, long after everyone knows the theory was soundly debunked. Also, pundits have taken liberty with a line from the first page of Genesis: “God said ‘let there be light’ and then there was light” by replacing the original Aramaic lettering of “let there be light” with Maxwell’s equations (which look every bit as mysterious to the layman as Aramaic).

1.3 Heinrich Hertz

In 1887, Heinrich Hertz was working on practical applications of Maxwell’s findings and demonstrated that under the proper conditions, inducing a spark on one side of a room could induce a simultaneous spark in a circle of wire with a narrow gap across the room. Hertz found that the intensity of the distant spark depended strongly on the size of the wire circle. He realized the loop size was “tuning” the receiving loop to match the frequencies emitted by the sender sparks, arguably the first demonstration of a radio transmitter and receiver.

For many years, the concept of frequency was referred to by the dimensions “cycles per second.” In the late 1960’s Hertz was honored for his work by renaming the unit of frequency as Hertz, (abbreviated Hz), where one Hertz is one cycle per second.

1.4 Electrical Resonance

While this course does not go into circuitry, one such aspect of electronic design will repeatedly occur, and that is resonant circuits. Let us address it early on.

A pendulum swings at a reliable rate, accurate enough that every grandfather clock uses a pendulum to track time. Many other forms of mechanical resonance are commonplace, especially in music: A piano key strikes the string, a church bell is hit with a hammer. When struck, these devices ring, or resonate, at well-defined frequencies, which in the case of musical instruments are known as notes.

In electronics, an inductor and a capacitor can be connected to provide an analogous electrical ringing resonance. Applying a narrow pulse will cause the circuit to ring like a bell, even though it cannot be heard directly. Actually, with properly sized components, including a loudspeaker, a ringing sound can be generated. However, audio frequencies are so low relative to radio frequencies that listening to them is impractical. Instead, test equipment called oscilloscopes shows instantaneous voltage in a circuit as a function of time. Viewing a narrow pulse applied to a resonant circuit on an oscilloscope will show the abrupt start of a sinusoidal vibration that slowly decays with time.

Applying a continuous signal of a single frequency to such a circuit will produce a large increase in amplitude only at the resonant frequency. This characteristic appears everywhere in radio.

Inductors are coils of wire that may be wound around a metal core. Inductors are used to convert energy between electrical currents and magnetic fields. Power utility transformers working at 60 Hz are built of huge inductors with thousands of turns of wire tightly wrapped around laminated iron cores. With increasing frequency, the appropriate size of inductors is reduced, until at 200 MHz a typical inductor may only be a few turns of wire an inch (2.5 cm) in diameter and equally long, with no core other than air. Values of inductors (or coils) use units called Henrys (after the early scientist Joseph Henry) or the letter H. However, one Henry is a large inductor, while at the frequencies and power levels of radio work, inductors are usually encountered with values stated in microhenries (10^{-6} Henry, or uH) or even nanohenries (10^{-9}H or nH). Inductors are symbolized in the literature by the letter L.

Capacitors at their simplest are two plates of metal separated by a nonconductor, used to store an electrostatic charge. They are symbolized by the letter C and their basic unit is the Farad (referenced by the letter F), but capacitors – much like inductors – usually are sized such that they are represented by microfarads (10^{-6} F or uF), and picofarads (10^{-12} F or pF; sometimes called uuf or “mickey mikes”). Note that nanofarads (10^{-9} F or nF) are rarely used, with values like 0.001 uF or 1000 pF substituting.

Combinations of inductors and capacitors allow energy to flow between electrostatic and magnetic fields in a manner analogous to the way a pendulum swing depends on transferring between potential energy and kinetic energy. This L-C resonance is used to build frequency selective networks, referred to as L-C networks, which are everywhere in radio.

2.0 Early Radio Receivers

Many experimenters began building on Hertz’s work; mostly in developing more sensitive receivers. An incredibly simple device is worthy of note – the crystal set. (To this day crystal sets are still available as kits, and children often find them interesting).

A crystal set uses a long wire as an antenna to pick up as much signal as possible, and couple it to a coil and capacitor set up as a L-C resonant circuit at a frequency matching a nearby transmitter. The coil is usually made of enamel-insulated (“magnet”) wire, with the enamel sanded off a strip so a movable arm can move a contact point along the coil to adjust the inductance and thus tune the frequency (along the top of the coil in Figure 3). A rectifying diode set up as a “peak detector” (see Figure 2) removes the radio frequency component, while maintaining the amplitude variations (“modulation envelope”), leaving a 120 Hz raspy buzz for spark signals (as spark transmitters ran directly off 60 Hz power), or the voice program for an AM signal. The first diodes were a crystal of galena ore and a short length of wire (“cat whisker”) that the operator would move around the crystal until he found a sensitive point. A pair of high-impedance headphones would turn the resulting signals into sound. (Note: modern low impedance headphones are not usable).

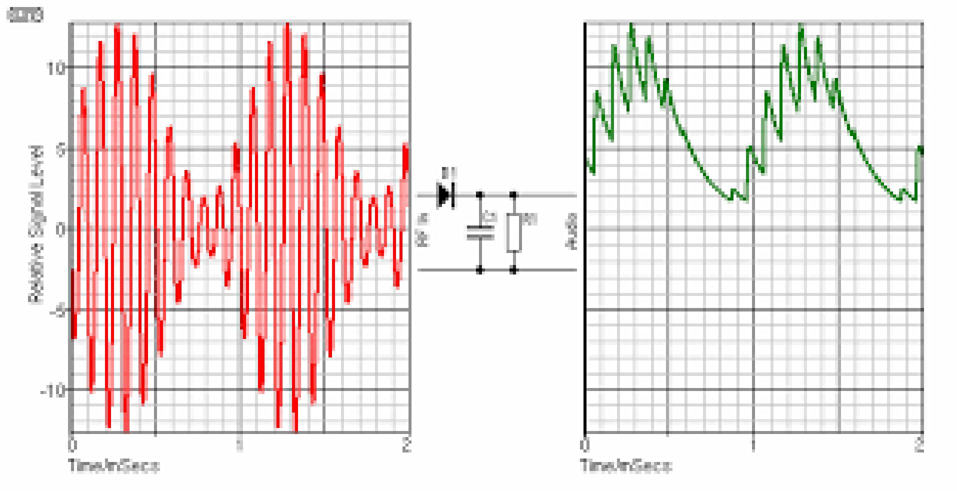

Figure 2 - Depiction of an AM waveform and diode demodulation as seen on an oscilloscope

Figure 3 - Crystal receiver (note galena crystal and "catwhisker" in middle)

2.1 Marconi and King Spark

In the early 1900s it was realized that recent work with sparks causing action at a distance led to the tantalizing idea of finally enabling communication between ships at sea. The existing Morse Code protocols and hand-keys long-used for telegraph messages lent themselves perfectly.

At first, the concept of long range communication was in doubt. Light only travels in a straight line; obviously other electromagnetic signals will do the same. It appeared at first that these spark-induced signals would only be good for short distances. However, the age of miracles was not over, and God has provided us a planet and atmosphere that can provide various forms of ducting and reflections to allow such signals to bend over the horizon. There will be more on this later.

Experiments progressed until Gugliemo Marconi was able to (barely) communicate across the Atlantic in 1902. Within a few years "wireless" equipment was installed on every ocean going vessel. A typical station used a multi-kilowatt spark generator that would hurl blue lightning when the operator pressed his Morse Code key (Figure 4). Huge sparks, and high voltages and currents on the hand key, made it a noisy and dangerous job (but the ultraviolet emissions did produce artificial suntans).

Sparks generate a broadband spectrum. A transmitter needed a tuning section with coils and capacitors to work with the antenna to select and transmit only the desired frequency. Similarly the receiver had to peak the sensitivity at the desired frequency where shipboard operators would all tune. These seemed to gravitate to the area around 250 kHz, which became a standard emergency frequency.

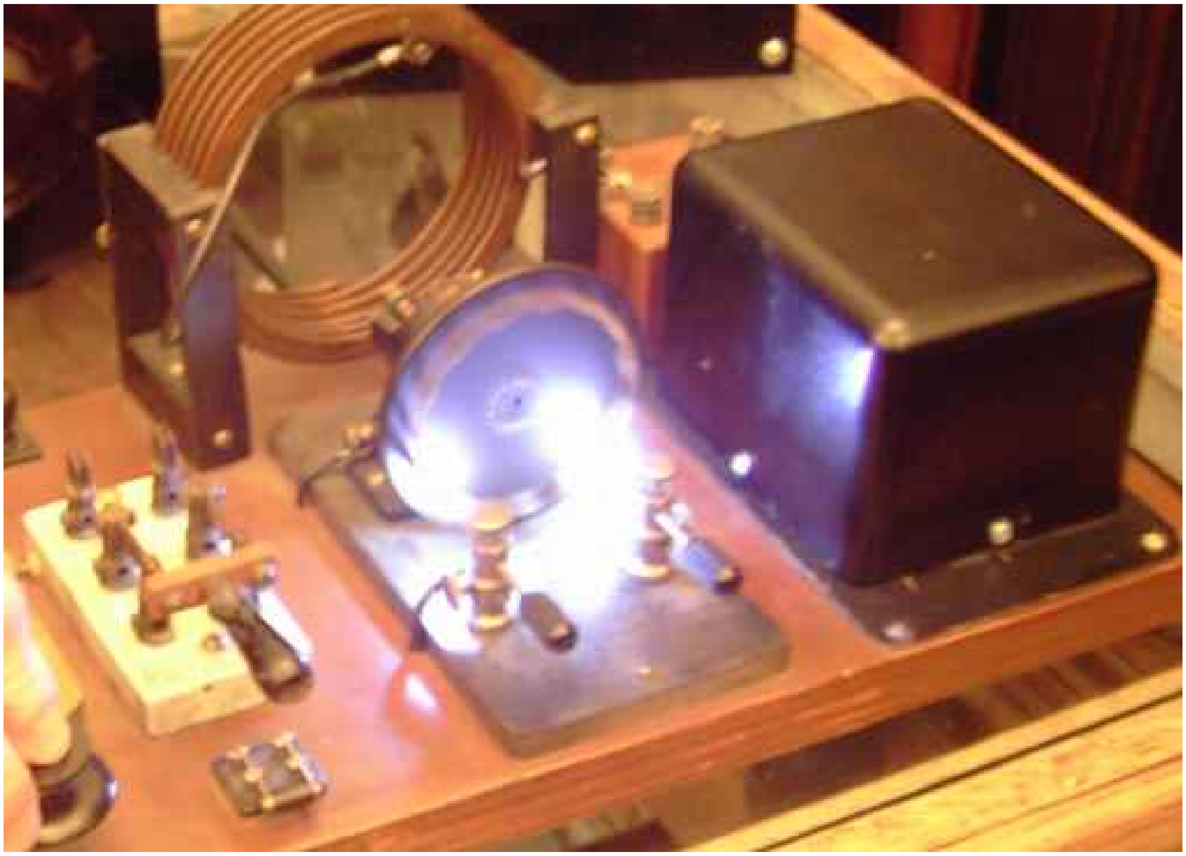

Figure 4 - Modern replica of a small spark gap transmitter in action (Imagine a big one)

The disastrous sinking of the RMS Titanic in April 1912 was the result of a long string of mistakes, arrogance, and just-plain stupidity. The only positive thing that came out of the entire affair was the reception of distress messages via wireless by the RMS Carpathia, whose Captain, crew, and passengers set the standard for emergency response from then on, efficiently rescuing survivors.

2.2 CW pushes out spark

Spark transmitters were not only power-hungry and dangerous to be around, but they were at their core broadband noise generators, spreading the energy over a wide swath of frequencies. The available filtering only narrowed the radiated spectrum somewhat, still taking up wide bandwidths that severely limited the number of usable channels. Researchers felt there had to be a better way, and with the advent of vacuum tubes, they developed what became known as Continuous Wave (CW) transmission. Since CW also uses Morse Code, which turns the signal on and off, the term “continuous” is a bit of a misnomer. It has been taken to mean when the signal is present (key down), it is a single continuous frequency with no other modulations.

The key part of a CW transmitter is an electronic oscillator that produces a single frequency. The invention of vacuum tube triodes that could provide amplification made the breakthrough possible. An oscillator is an amplifier with a circuit that provides strong, positive feedback at a single, well defined frequency. (A screeching public address system is the same thing – an amplifier with feedback driving it to oscillate at a single frequency). Additional vacuum tubes amplify the signal to suitable levels for transmitting.

There are two basic types of oscillators: L-C tuned and crystal controlled. An L-C tuned oscillator uses the values of the inductor (L) and capacitor (C) to determine the frequency. L-C oscillators are simple and reliable, but the frequency is affected by many variables including temperature. (Vacuum tubes have internal heater filaments that literally warm up with time, and a transmitter may not stabilize for hours).

Quartz crystals can be cut and polished to resonate at a specific, precise, repeatable and stable frequency, much more so than an L-C oscillator. However, they are not tunable. The frequency of a crystal oscillator can be ever-so-slightly adjusted to put it exactly on the proper frequency, but nothing more.

2.3 The early CW receiver

Like a CW transmitter, the matching receiver now had amplification to pick signals out of the Aether. An early radio receiver design is the Tuned Radio frequency (TRF) receiver, with a series of amplifier stages each with L-C tuning to bring signals at the correct frequency up to the necessary level to be demodulated. The string of tuned amplifiers would provide a narrower reception band than a single tuned stage; the receiver could optimize reception of the narrowband signals. After demodulation of the audio, additional amplifiers would further increase the audio to the level that it could drive a loudspeaker.

Figure 4 shows a TRF receiver block diagram. However, in certain specific applications (single frequency, very close range) the TRF design is still usable and the diagram shown is of a modern receiver. Note the input L-C filter, the sequence of RF amplifiers and a detector.

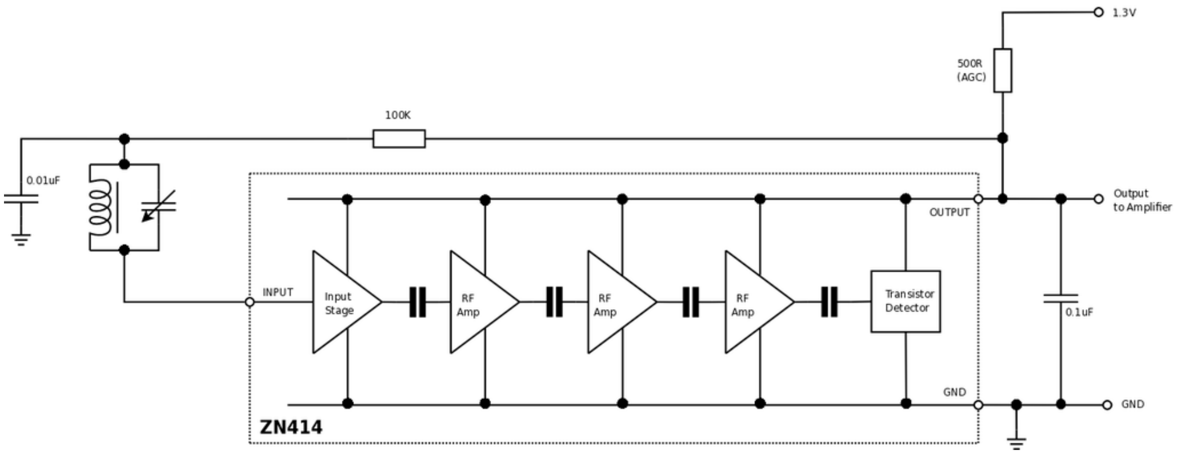

Figure 5 - Tuned Radio Frequency (TRF) receiver

The series of amplifiers gives a TRF receiver its signal amplification. However, each stage has to be individually tuned to the proper resonant frequency by an L-C circuit. To change frequency, each stage must be re-tuned. It is not a simple task to maintain the proper bandwidth provided by the combined (“cascaded”) filters.

Where a simple receiver could demodulate the rapidly varying amplitudes inherent in a spark signal to output the characteristic raspy sparking sound, CW needed a more complex detector.

The regenerative detector is similar to another tuned amplification stage, but with an adjustable feedback loop coupling a small portion of the output signal back to the input, usually by adding a smaller “tickler” coil around the coil in the L-C circuit. The operator would “ride” the regeneration control to keep it just on the border of breaking into oscillation, and an incoming signal would tip the balance to set it oscillating, generating a tone in the headphones. When properly set, the oscillation stops when the input signal does.

While spark transmissions occupy a part of the spectrum, a CW signal is by its nature very narrow around the center frequency, so the number of usable channels open up tremendously.

The increased receiver sensitivity and narrower bandwidths improved communication range significantly. The ability of the radio signals to bend with the curvature of the earth continued as it did with spark, but the narrower signals and high amplification receivers made regular communication over moderate ranges more reliable.

3.0 Radio Reaches Every Home – The AM Broadcast Band

Figure 6 - An AM console receiver of the 1930's

3.1 The early AM receiver

It did not take long until CW signals began to vary amplitude with voice or even music, opening up an entirely new field of Amplitude Modulated (AM) broadcasting. Frequencies were soon assigned for an AM broadcast band from 550 to 1650 kHz, where they remain to this day.

Referring back to Figure 2 shows a representative waveform of an AM signal showing a single modulating frequency (whistling into a microphone will provide a fairly good 1 kHz sinewave). Filling in the area between the upper and lower sinusoids is a much higher sinusoid at the radio frequency

The simple detectors used for spark, with the galena ore replaced with a diode (vacuum tube diodes at first, and later germanium) worked very well for AM, so the regenerative detector was not necessary. However, the regenerative detector can be readjusted to detect the amplitude variations of an AM signal while providing some moderate signal amplification, so some dual-purpose CW - AM receivers became available.

One interesting aspect of the AM radio band is that a tremendous amount of signal amplification is not needed. Our planet is one tremendous spark transmitter, with every thunderstorm adding to the din. Enough of the lightning-induced emissions travel far enough to establish a "noise floor" of atmospheric noise that covers up weak signals. To be heard over the din, AM broadcast transmitters must run high power to be heard over their listening area (50 KW is common). This further helps keep the costs of the home receiver down.

The long-distance effects that the old wireless operators depended continued into the AM band, which is fairly well controlled during the daylight hours. At night, “ground-wave” causes signals to better follow the curvature of the earth. Under ideal conditions, on a clear cold night, AM stations over 1000 miles away (1600km) can come in every bit as well as the local ones.

3.2 The Superhet Receiver

Figure 7 - AM vacuum tube table radio

With a huge new AM broadcast consumer market, the TRF receiver was becoming a problem, needing a careful re-tuning of several stages each time the listener wanted to change the station. A radio receiver that needs a trained operator was holding down consumer acceptance. Something that anyone could operate was needed, and quickly.

The superheterodyne receiver design adds complexity, but makes the receiver much easier to use. The home radio listener no longer needs to regularly retune multiple stages. The idea is to build a single-frequency receiver at what is called an intermediate frequency (IF), and factory-tune it to get the desired performance that will hold constant for the life of the receiver. Then, some device was needed up-front to allow the user to select a desired frequency and convert that signal to the IF frequency. There are two new features introduced by the superheterodyne: the IF amplifier, and the converter that changes a selectable input frequency into the constant IF frequency.

3.2.1 The IF strip

The single frequency IF receiver typically operates 455 kHz in most AM broadcast band home radio receivers. The “IF strip” consists of two or three L-C tuned amplifier stages that are factory-adjusted not only to be right on frequency, but also to have the optimum bandwidth for reception of AM broadcasts (approximately 15 kHz). Each stage has several adjustments, usually implemented with screw-adjustable metal cores in the inductors. Cores increase the inductance of a coil, and add adjustability as they can be moved in and out. They are often made of ferrite, which is a ceramic-like material made with powdered iron-based alloys. Because inductors convert currents into magnetic fields, interaction between stages is avoided by encasing them in small metal shields, resulting in the name “IF cans.” Typically there is one IF can per stage and they usually include the resonating capacitors so as to allow factory pre-tuning of the can before installation.

3.2.2 Frequency converter

The frequency converter is the second section added to make a superhet receiver. Since the selected input signal is to be converted to the IF frequency, the converter must be tunable. There are two parts to the converter: the mixer and the local oscillator (LO).

3.2.2.1 Mixer

The key to the process is the function known as heterodyning which is based on a mathematical process. If, in a mathematical exercise, two sinewaves of different frequencies are multiplied, when all the resulting terms are simplified, the resulting terms are primarily a sinewave at a frequency that is the sum of the two input frequencies, and a similar sinewave at the difference frequency.

It is not necessary for the conversion stage, or “mixer” to exhibit a true multiplication; an amplifier with any nonlinearity will generate the necessary sum and difference frequencies.

This is not as abstract as it first seems. Flying on an airliner, sometimes the droning sound of the engines will begin to show some slow wobbling variations. Or, tuning a musical instrument against a tuning fork, sometimes creates a wobbling characteristic in the ears, slowing down as the instrument is brought in perfect tune. These effects are known as “beat notes” and are generated in our ears due to their logarithmic characteristics. What we hear is a sound component at the difference frequency between the two engines or two separate musical tones. There is also a sum frequency, but the wobbling characteristic of the difference really stands out.

3.2.2.2 The Local Oscillator (“LO”)

The second signal used in the heterodyning process is generated in a “Local Oscillator” (LO), a circuit that (similar to part of a transmitter) generates a frequency dependent on its L-C network. The LO must be tuned to select the desired station. For a receiver with an IF of 455 kHz, and, let’s say, 1 MHz as the local oscillator frequency, inputs at either 1.455 MHz (sum) or 545 kHz (difference) will send IF frequency signals to be processed. One is the desired frequency; the other, referred to as “image,” must be rejected. Note that the two are identical, and neither is fundamentally good or bad; it all depends on user preference.

3.2.2.3 The Preselector

It is customary to add another tunable L-C circuit to the input of the mixer to select the desired choice of the two possible inputs: One is selected and the other is rejected (preselection). Note that we now have two tuning adjustments to change the station: local oscillator frequency and preselector tuning. Customarily, both L-C circuits use variable capacitors that are “ganged” on the same shaft. Much effort goes into the design of such combined units so the two track closely over the AM broadcast band.

When the same ferrite core is used for both the preselector and local oscillator coils, any interaction between the two is not an issue, as they are supposed to be combined anyway. The use of such "ferrite loopsticks" provides enough pickup of the received signals into the preselector that an outdoor antenna is not needed for strong signals, tremendously improving convenience of use.

Figure 8 shows the block diagram of a superhet receiver and Figure 9 shows the rear view of the chassis of the table radio of Figure 7. Note that the preselector is referred to as "RF amplifier", which simply means the preselector amplifies the input signal.

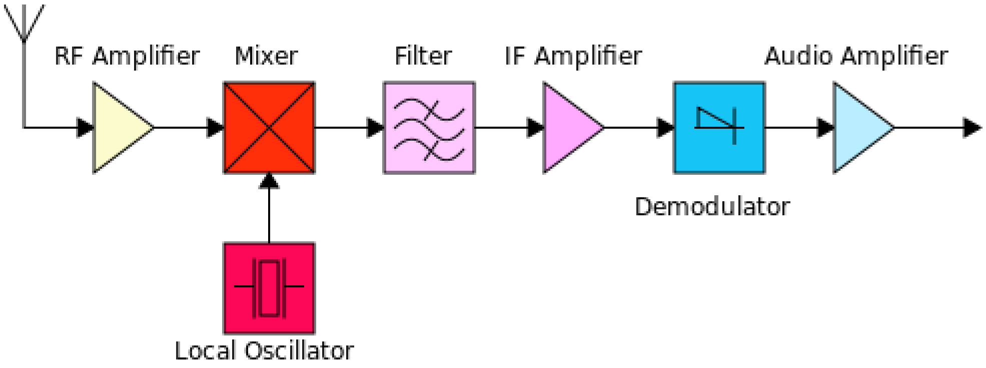

Figure 8 - Superheterodyne block diagram

Figure 9 - Internal view of a "5 tube" superhet" table radio of Fig 7. Notice ganged variable capacitor towards the left, and the IF can towards the right.

3.2.3 Technical Improvements

The AM broadcast band is still widely used today and many evolutionary improvements have been made over the years. Here we will take a small detour from the chronological flow of increasing frequency to dwell on more recent developments in AM radio technology. Such developments are not limited solely to the AM band; many also have applications at higher frequencies.

The development and popularity of car radios led to a new problem. In a moving car, the received signal will vary as the signal path between the transmitter and the car is occasionally blocked, or even reflects off nearby structures. The solution is an automatic gain control (AGC) feedback loop to reduce that effect. A circuit monitors and averages the audio output level, and develops a voltage indicative of average output. The voltage is sent back to one of the amplifier stages, to reduce its amplification as necessary to keep the audio level constant despite all the changes as the car moves.

Another necessary automotive change was pushbutton tuning. The term “distracted driving” didn’t exist then, but manually tuning in a radio station for best clarity while driving probably is even worse than texting. A pushbutton radio has a set of tunable inductors that are selected by pushbutton switches. Each inductor is tuned to the desired station – hopefully when parked or by a passenger – and holds that inductance indefinitely.

With time, AM broadcast radios became smaller and more convenient, although the overall philosophy has remained the same. At first, miniature vacuum tubes made the radios smaller, and some awkward so-called “portables” resulted (although the high voltage “B” batteries were a nightmare). The transistor absolutely revolutionized the portable radio to the point that table radios became a rarity. Car radios became much more user friendly, as transistors run directly from the automotive 12 volt electrical systems, without the need to generate high-voltages for vacuum tubes.

In recent years, designs have focused on using integrated circuit technology and automated assembly to reduce costs. The details have changed, but they all link back to the original designs.

3.3 The 3 – 30 MHz HF band and Shortwave Radio

In parallel with the maturation of the AM broadcast band in the 1920’s, continuing work was being done on long-distance communication. As vacuum tubes evolved, higher frequencies became possible, but no one went there because God had not provided any of those wonderful atmospheric propagation effects way up there in the great unknown. Meanwhile, the cost of experimenting had dropped to the point that many interested hobbyists began investigating for their own curiosity and amusement. The government decided to keep them away from its important work and allowed those annoying amateurs (who for some reason called themselves “hams”) to only use the vast wasteland above the AM broadcast band.

The amateurs played around and discovered just how wrong the conventional wisdom was; The 3 – 30 MHz segment of the spectrum simply has the most intriguing atmospheric effects for long-range - even international propagation. To this day, the HF band is the premier way to communicate over very long distances without any infrastructure; Todays radio hams still assist in emergencies by using HF to communicate to the rest of the world when the usual methods of communication are down.

3.3.1 HF propagation

At frequencies below 10 MHz, groundwave is the most common form of long-range communication, as the signals can bend and follow the surface of the earth. This effect is usually weaker with sunlight and best on a cold clear night, but it can be unpredictable.

The upper levels of the atmosphere are subject to molecular ionization from bombardment of all the rays and particles coming from the sun. Of course, the effects are greatest in daylight, but sometimes the ionization can last well into the evening. From 10 to 30 MHz the primary form is signals reflecting off the aptly named ionosphere. Signals can even bounce a few times between the ionosphere and the earth's surface. Propagation to the other side of the earth is not uncommon in that mode.

HF propagation can be infuriatingly variable. It varies with almost anything: solar radiation, weather, season, but primarily with sunspots. Long ago, astronomers observed that sunspots waxed and waned on an 11 year cycle, but it wasn't until HF propagation was studied that it was realized that the particles emitted from the sun in sunspots raise the ionization level of the earth's atmosphere, and thus its reflectivity. During sunspot cycle peaks, HF propagation can become quite predictable.

Another interesting aspect of the HF band is that as the frequency goes up, the energy content in electrical sparking (regardless of source) goes down, lowering the atmospheric noise floor. The drop off in the energy content of spark transmissions at the higher frequencies may well have explained the early bias against higher frequencies

The key point, however, is that elimination of atmospheric noise allow receivers to be made more sensitive to achieve longer range, even with lower transmitter power. It is not a free ride, however, as there still is a limiting noise floor, although it is much lower. (This is discussed in the FM band section 4.1.)

3.3.2 HF shortwave radio

Since the 1950's many nations have sponsored "shortwave radio" international AM broadcast stations in the region of 5 to 10 MHz to communicate their ideas to people in other countries. Much effort and expense has gone into such stations as Voice of America, BBC, and Radio Moscow, but many other nations have their own flagship stations. The better funded stations actively utilize the propagation available at any moment, using antennas that can direct the main part of the signal at a "launch angle" to best match the atmospheric effects present at the time.

3.3.3 HF Communication Receivers

Shortwave receivers are very much like AM broadcast receivers, but they usually cover the entire 3 to 30 MHz HF band, and the wider frequency range brings the need for some changes. The obvious first issue was that the HF band covers a 10:1 frequency range and practical L-C tuning components cannot cover that wide a range. So the coverage is broken down into smaller" bands," with different L-C circuits switched in as necessary. As a result, shortwave receivers are notable for their complex bandswitching, with impressive arrays of isolated switch sections running through the entire receiver.

3.3.3.1 RF amplifier stage

It was mentioned earlier that amplification of the incoming signal in an AM band receiver is optional due to the high level of atmospheric noise covering weak signals. The reduction of atmospheric noise in the higher frequencies of the HF band exposes those weaker signals, driving receivers to have higher amplification, to provide more sensitivity and correspondingly longer distance communication. One obvious step up is adding a tuned amplifier to the preselector as a "Radio Frequency (RF)" amplifier.

3.3.3.2 Beat Frequency Oscillator (BFO) or second mixer

Even as the AM broadcast band occupied the public imagination, CW point-to-point communications continued for ships, airplanes, etc. However, simple AM detectors do not work on CW, and a new detector was developed.

As described earlier, a mixer and local oscillator are used to change the input radio frequency (RF) to the intermediate frequency (IF). A second mixer and LO are now placed at the IF output to heterodyne the IF down to an audible frequency (about 1 kHz). Again, the mixer is ideally a true multiplier, but in reality something simpler will suffice. In this case, the local oscillator is called the beat frequency oscillator (BFO) and it typically operates about 800 Hz above or below the IF frequency so the mixer will accept a sinusoidal signal of a 455 kHz CW from the IF, signal, and heterodyne it to an audio tone. The exact BFO frequency is usually adjustable over a narrow range to the operator's preference. (Never underestimate the frequency filtering capability of a well-trained ear).

Note that the image frequency is so close that a normal L-C filter cannot separate the two, and a signal will come in equally well on both sides of "zero beat" (the null point where the two frequencies are close enough that the beat note becomes sub-audible.

This was recognized as an issue for crowded frequencies, where the receiver bandwidth is much wider than the signals passing through it. This led to more developments.

3.3.3.3 Updating IF frequencies and filters

Not only were redundant CW signals at issue, but there were other limitations on the 455 kHz IF strip used on AM broadcast receivers.

The AM broadcast band covers nominally 0.5 to 1.5 MHz, so an IF frequency of 455 kHz has the desired and image frequency separated by 910 kHz. A simple L-C filter easily provides the necessary separation. As the input frequencies went up, the percent differences between the desired and image signals kept getting smaller (and L-C filters work on percentages), so the undesired signals were now bleeding through.

A move to new kinds of IF filters produced a number of possibilities. There is nothing sacred about any IF frequency, and quartz crystals were available with highly resonant frequencies in the range of 3 to 10 MHz. A crystal could be placed between IF stages with a bandwidth much narrower than any L-C filter. Properly set, such a filter can pass a bandwidth of 500 to1000 Hz; narrow enough to filter out the other side of zero beat. Of course, such a filter is unacceptably narrow for voice, but it filters out a lot of extraneous clutter and noise. Shortwave receivers began to have higher IF frequencies (4.5 and 9 MHz were and are popular), with a switchable crystal filter; on for CW, and bypassed for AM.

For AM reception, peak detectors were still needed, and new ceramic 455 kHz filters were developed with the perfect bandwidth for voice signals, plus they were smaller and cheaper than IF cans, as well as adjustment free.

3.3.3.4 Double and triple conversion

An alternate approach to high frequency AM reception uses a third heterodyning stage, with its own mixer and LO to have a high IF and a low IF. This double conversion technique allows the use of the ubiquitous 455 kHz ceramic filters in the low IF, while the high IF has its own L-C preslector to work in conjunction with the RF preselector to prevent IF images. Some manufacturers carried the system to the next step, with triple conversion receivers. Keep in mind that such multiple conversions have signals at multiple frequencies, and careful design is necessary to avoid any of the possible mathematical combinations of those frequencies from finding each other.

More recently, some new filters have been developed at quite high frequencies (70 MHz seems to be a standard) and some receivers have been built with an IF that is actually higher than any of the input frequencies. This does provide some advantage in that the heterodyning images are way out of any tuning range.

3.3.3.5 Stable oscillators

By this point the technology had improved to the point that an old problem could no longer be ignored. The components of an L-C circuit vary with temperature, humidity, age, etc. all of which lead to slight drifting of the oscillator frequency. (L-C filter bandwidths change slightly too, but the errors are not as noticeable). At first, such drifts were too small to notice, but as the communications systems improved, the drift was becoming a problem.

The same quartz crystals used in the IF filters can also be used in oscillators instead of L-C circuits, with stabilities of under 100 parts-per-million (ppm), and with some effort can be under 10 ppm. The downside is that crystals are only tunable over a range of under 1000 ppm, which is usable for a BFO, which only tweaks the CW audio tone. Typically a modern receiver has a small variable capacitor to tweak the crystal BFO, referred to by any one of several names: fine tuning, clarifier, or receiver incremental tuning.

Double conversion receivers often use several bandswitch-selected crystals in the high frequency LO, one for each sub-band, while the mixer and any RF amp are pre-tuned to pass a moderate sub-band of frequencies at RF and then convert them down to the first

IF, which is wide enough to pass the entire sub-band without retuning. The second local oscillator (for the mixer converting from high IF to low IF) is usually where the main tuning control is located to select individual signals for the low IF from the selected sub-band. This second LO can take a number of forms, all of which have stability as their main criteria.

The classic form of the second local oscillator is still L-C based, but with high priced, stable components to improve stability. A typical example changes the tuning mechanism from the classic variable capacitor to a large "permeability tuned" coil, where the inductor is on a tubular form with a core that is moved in and out of the coil by the tuning control to provide a fairly linear tuning rate. The coil is resonated by fixed-value capacitors that have well defined temperature coefficients, both positive and negative, and with proper selection will combine to ideally cancel out the temperature coefficient of the coil.

3.3.4 Single Sideband

For some years, theorists had been predicting a new form of voice modulation, derived from AM but quite different in use. It was referred to as Single Sideband (SSB), and now the equipment finally had the stability to make it practical.

AM radio may be easy to receive, but it has a large cost in performance that becomes obvious upon observing the transmitted waveform on suitable equipment. If the operator whistles into the microphone of an AM transmitter monitored by an oscilloscope, which shows instantaneous voltage as a function of time, you will see a sinewave at the radio carrier frequency, with the amplitude much more slowly varying somewhat with the modulating frequency; roughly 1 kHz in the case of a whistle. Refer back to Figure 2, or note Figure 10 which is a 1964 commemorative postage stamp showing an AM waveform:

Figure 10 - AM waveform on scope

However, viewing the exact same signal on a spectrum analyzer, which instead shows the instantaneous power as a function of frequency (as opposed to time), you will instead see three vertical lines as shown in Figure 11.

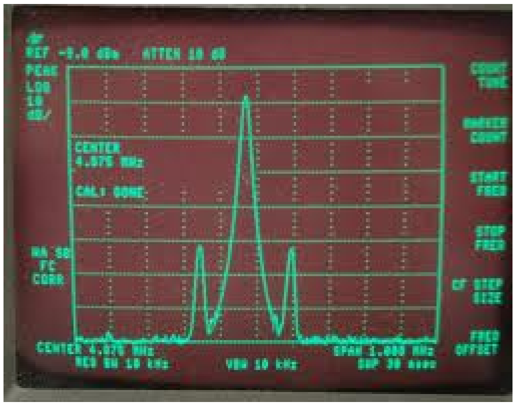

Figure 11 - Spectrum analyzer display of AM signal

The large, central line is the “carrier” which is exactly what would be seen for an unmodulated CW signal at that frequency. On either side of the carrier, separated by the modulating frequency (in the case of a whistle, 1 kHz), there are two smaller vertical lines. These lines contain the information being transmitted, and in the case of voice or music, they will move around as the information changes, while the carrier sits there rock-solid, even for those moments when there is no program material and the sideband amplitudes go to zero. Not only is all of the useful information in the sidebands, but the two sidebands are redundant mirror-images.

Eliminating the carrier and one sideband provides a huge increase in effective power. To reconstitute the original audio, however, the receiver must restore the carrier at exactly the right frequency. However, doing all of this requires levels of precision and linearity that simply were not practical until the 1950’s, when Single Sideband was ready-to-go.

Generating a SSB signal is done in a process resembling running a superheterodyne receiver in reverse, where a local oscillator signal is multiplied with the microphone signal to generate an intermediate frequency (IF), which is then heterodyned up to the transmit frequency.

All of the mixers mentioned so far have been able to approximate a multiplication of two frequencies to get the sum and difference frequencies by utilizing nonlinearities. Creating or decoding SSB signals, however, requires true multiplication, with a linear response over a wide range of amplitude of the voice modulation. A proper multiplier is called a “balanced modulator,” and it multiplies the oscillator and microphone signals to generate two sideband frequencies (sum and difference of the inputs). However, true multiplication does not pass through any of the input carrier signals (other than a small bit of leakage).

With the carrier is gone, now one sideband must be selected and the other rejected. (While there are standards on when to use which sideband, they are not for any technical reasons).

Quality filters can separate the two sidebands at the IF frequency. One requirement of such a filter is that it be of the specific width to pass a suitable voice spectrum (300 to 3000 Hz), but at the IF frequency. A filter must pass the desired IF signals over a 2700 Hz channel undisturbed, yet reject the unwanted sideband 600 Hz away, as well as what little is left of the carrier only 300 Hz away. Obviously, the filter needs a sharp cutoff characteristic. Originally, there were mechanical filters that worked well, but were at such low frequencies that receiver-style double conversion techniques in reverse were needed for transmit. Later on, filters using a number of crystals at slightly different frequencies were developed to provide the perfect bandwidth for SSB. Their high operating frequency (9 MHz is common) allowed for single up-conversions to the transmit frequency.

Demodulating the signal in the receiver reverses the process used to generate it. First, the signal must be passed through a filter of the same characteristics as the transmit filter (some designs use the exact same filter both ways). Then the original (long since discarded) carrier must be replaced precisely. The IF signal is linearly multiplied with a locally generated LO signal, reversing the process where the signal is generated. When a balanced mixer is used to change the IF back to audio, it is called a product detector, which is backwards compatible to also demodulate CW. The BFO must be very stable to properly recreate the carrier frequency accurately enough, and such oscillators are often voltage controlled crystal oscillators (VCXO) that allow very slight variations of an accurate reference to allow the operator to fine-tune for best audio quality.

In SSB the transmitter shares much receiver circuitry, only in reverse. It has become common to combine both the transmit and receive functions in one unit (transceivers). In such a case, the receive/transmit switching can be pervasive, and designs devote much hardware to the receive/transmit switching and the (previously mentioned) band switching.

Re-inserting a carrier does not guarantee that it is close enough in frequency to re-create the original voice. In a lab, all the frequencies can be lined up to regenerate audio, but real-world radio propagation adds enough variations to produce noticeable voice degradation. Manual adjustment of the demodulator oscillator frequency is often necessary. One of the names for the BFO tuning control is “clarifier” and that comes from SSB use where it is used to fine tune the original audio. Properly tuning in an SSB signal requires listening to the received audio, which is said to sound like the quacking quality of the voice of the cartoon character, Donald Duck. As the oscillator is tuned to bring the pitch into normal voice frequency range, the audio becomes comprehensible. However, SSB audio is never of good fidelity, so it is strictly limited to communication-quality voice. However, for that price it provides a tremendous increase in “talk power” and/or reduction in transmitter power consumption.

Everyone has heard Neil Armstrong's "…a giant leap for mankind…" transmitted from the moon. Listen again and you will hear the slight "quacking" characteristic unique to SSB in his voice. Thinking about it, there was a specific frequency assignment, no need for high fidelity, and a desperate need for reliable communications with the absolute minimum power consumption. SSB was the obvious choice

3.3.5 Recent developments

As we did earlier, we again take a small detour from the chronological flow of increasing frequency. Recent developments in VHF and UHF receivers also have applications in HF receivers, so they are introduced here and referenced later.

3.3.5.1 Phase Lock Loop (PLL)

The phase-lock-loop is the next step. Our discussion will first go into the building blocks before they are interconnected. There are two oscillators, the reference oscillator (F_1), and the Voltage Controlled Oscillator (VCO), each with its own frequency divider, a frequency/phase detector and a filter. See Figure 12.

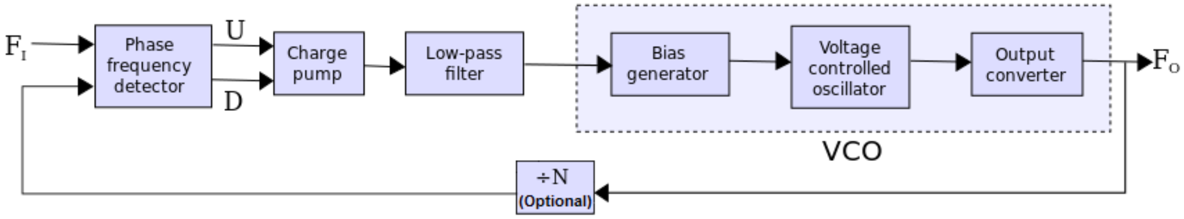

Figure 12 - Phase Lock Loop block diagram

The frequency reference is a crystal controlled oscillator, designed for maximum frequency stability, even possibly to the point of mounting it in a small "oven" to always maintain it at the exact same temperature. Reference oscillators often output standard frequencies, such as 10.000 MHz. or some similar round number.

With the advent of digital ICs, inexpensive frequency divider circuits became available that produce an output at a well-defined fraction of the input frequency, yet with the same accuracy and stability of the source. Note that the division ratio must be an integer, and the reference oscillator is typically divided down to frequency under a few kHz.

The second oscillator is an L-C type, with a variable tuning capacitor that rather than being manually tuned, changes capacity with applying a variable voltage to a back-biased diode. (Back-biasing a diode blocks any current flow.) Such "varicap" diodes," optimized for this purpose, can have impressive voltage-to-capacitance ranges.

The VCO is brought to another digital counter, but this one divides by a variable number that is set to determine the desired output frequency.

A frequency/phase detector accepts the two digitally divided oscillator signals and provides an output that varies with their frequencies. If one frequency is higher than the other, the output goes to its maximum positive voltage; reversing the inputs drives the output to maximum negative. If the two frequencies are the same, the output varies as a function of the phase between them.

The filter averages the output of the phase detector and connects it to the tuning input of the VCO in such a way that it closes a feedback loop to drive the divided VCO frequency and phase to match that of the divided reference input.

If conditions allow the loop to close, it achieves “phase lock,” where the two inputs to the frequency/phase detector are identical. At this point, the raw VCO frequency is equal to the divided reference frequency multiplied by the VCO division ratio.

Consider a reference oscillator at precisely 10 MHz, divided by 100,000 to get 100 Hz. If the VCO divider is set to 75,432, to achieve phase lock, the VCO must be operating at 75,432 times 100 Hz, or 7.5432 MHz. Note that if the VCO divider changes by one count, the loop will drive to change the VCO by 100 Hz.

When used as a receiver local oscillator, the VCO raw output is used as the LO signal to the receiver. In this, example changing the VCO divider settings drives the loop to change the VCO frequency in 100 Hz steps (just so long as it is within the tuning range of the VCO).

Of course, a phase lock loop is a classical control system and subject to all the things that can go wrong with any control system, but properly designed, it is a tremendous boon to receiver stability.

One weakness with phase lock loops is that the loop is only updated at the divided reference frequency, and it can drift between samples, causing an effect known as phase noise. The lower the sample frequency, and correspondingly the smaller the tuning steps, the more stable the VCO needs to be when free-running on its own. Often a VCO is a high-quality, stable L-C oscillator, to keep the phase noise down. With proper attention to detail, receiver tuning with phase lock loops can be controlled down to 10 Hz steps.

3.3.5.2 Direct Digital Synthesis (DDS)

Working with phase lock loops rapidly shows a problem in that frequencies can only be generated at multiples of the divided reference frequency. If fine tuning steps are required, the phase noise becomes an issue. What is needed is dividers that can divide by a fractional number.

Direct Digital Synthesis achieves the long sought fractional division ratios with computer technology. A high precision crystal oscillator generates a clock appreciably higher than the maximum expected output frequency. A large counter is called the phase accumulator. Every time a clock pulse is generated, a preset number is added to the accumulator. The preset number is chosen such that the phase accumulator counts up to full count, resets to zero, and restarts for every cycle of the desired output frequency. The preset numbers adding into the phase accumulator do not necessarily have to always be the same on each count, so fractional adjustments are possible.

As the counter cycles through its paces, the highest weighted bits are output. If those bits were converted to an analog voltage, it would be a triangle wave, ramping up to a peak as the phase accumulator builds, and then suddenly dropping back to zero as the phase accumulator resets. However, the digital count is first passed through a lookup table before the analog conversion so as to provide instead a sinusoid. This technique causes the output sinusoid to have some rapid, small variations. However, passing the resulting signal through a narrow L-C filter, averages out the variations to provide the desired frequency.

Both the PLL and the DDS create signals with a stability derived from a single crystal controlled clock. As computer technology improves, however, the DDS has the greater potential for improvement and much new design is focused there.

3.4 Signal security; Encryption

The military obviously dislikes having their communications open to anyone with a receiver, and have always wanted to encrypt their communications. In World War II, when CW was the primary method of radio communication, various codes were used, with the operators encoding and decoding messages as they went. As the technology moved on, the operator was freed from actually doing the encryption.

3.4.1 Frequency hopping radios

Historically, plans were distributed to radio operators to manually change to a different frequency periodically. With the advent of phase lock loops for transmitter and receiver local oscillators, it was a logical step to let the receiver do it automatically to a predetermined code that defined the periodic changes. The signal sources must change frequency at the exact right time, often several times a second, to avoid gaps in the speech. The codes are loaded into the receivers and transmitters at the appropriate time and work automatically from there.

When a receiver first receives a coded signal, it cannot immediately follow the frequency hopping code. The standard approach to acquire the code is to have one frequency that is only used at one place in a repeating cycle. The receiver is tuned to that frequency and waits until a signal is received, at which point it counts off the time on-channel and then jumps to the next frequency in the code. From that point on, the receiver should be synchronized.

3.4.2 Encrypted radios

Other forms of encryption exist. Vocoders are stand-alone boxes that digitize the operator’s voice and use a code to change it. It is then turned back into a voice signal for transmission, reception, and regeneration in an equivalent unit.

Spectrum spreading is a technique where a transmitter adds high frequency components to a signal to intentionally make its spectrum much wider, but the HF band does not usually allow this, so it is reserved for the UHF and higher frequencies.

3.5 Software defined radios

The big thing in recent years is software defined radios. Computer technology has come to the point where the IF frequency of a receiver can be digitized and brought into a computer, where filtering, detection, and all the other parts of signal reception can be done in software. The software can have an entire library of say, IF filter characteristics, so the operator can choose the most appropriate one for any particular situation; a capability unaffordable in hardware. However, a receiver that does only one job well is still often the best way to go. Keep in mind that designing a software defined radio requires full knowledge of legacy receiver design, as changing from hardware to software does not alter the fundamentals.

As A/D conversion gets faster and more sensitive, digitization is slowly creeping ever closer to the antenna jack. It will be interesting to see how close it actually gets.

4.0 VHF and UHF Communication (30 to 400 MHz or so)

At frequencies above 30 MHz, the atmospheric noise level due to worldwide lightning is just about gone, but the magic of HF propagation is also pretty much gone too, and things mostly revert to the line of sight. There still can be rare atmospheric propagation events, but they cannot be depended upon.

At higher frequencies, there is also more spectrum for wider signal bandwidths, and accompanying higher information rates.

The earliest VHF AM receiver was the superregenerative design, which like the regenerative receiver used on HF, is based on an amplifier operating on the verge of oscillation. The superregenerative receiver builds up towards full oscillation, and is then reset, or “quenched” to re-start again. The oscillation builds faster with stronger input signals, and reaches a higher point before it is quenched. It delivers incredibly high sensitivity from only one vacuum tube (or more recently transistor). However, there is only one L-C circuit used which provides little filtering, and the weakness of the design is that it cannot be made narrowband. If bandwidth is not an issue, nothing gives more “bang for the buck”

The commercial VHF Air Traffic Control band at 118 to 136 MHz and the military UHF Air Traffic Control band at 225 to 400 MHz are exceptions, being classic, narrow band, AM systems. In this case, being limited to line-of-sight is a plus, as extreme range is not required, and channels are reused at intervals across the country.

Such AM receivers are quite similar to the HF band receivers already discussed except for the higher frequencies. One requirement is to have a higher first IF to avoid mixer images. Since they only use AM modulation, CW and SSB capabilities are not needed. Standardized Air Traffic Control Frequencies are the rule, and receivers operate on well-defined channels.

Military aircraft need to have both VHF commercial and UHF military receivers, so as to be good citizens while transiting civil airspace. The UHF receivers, however, often have voice encryption to keep transmitted commands secure in combat.

4.1 FM Broadcast

More typical of VHF is the FM broadcast band of 88 to 108 MHz, where wider bandwidth provides higher fidelity transmissions. In addition, the broad bandwidth of FM broadcast virtually eliminates the remaining atmospheric noise from local lightning storms. FM signals are by their nature always at the same power, but the modulation varies the frequency just as AM varies amplitude. The FM band allows 200 kHz wide signals (as opposed to 10 kHz for AM). In FM broadcast, the frequency varies within a specific "deviation" range, which in turn varies with the voice or music signal frequency. In FM the actual bandwidth is the sum of twice the deviation plus twice the instantaneous modulation frequency. The resulting spectrum on a spectrum analyzer display is much more complex than an AM signal, even for single frequency modulation.

FM receivers need to have an IF bandwidth of approximately 200 kHz, typically at a standard 10.7 MHz frequency, where suitable filters are readily available. Since a proper input signal is always of constant amplitude, a high gain amplifier and a limiter keep the signal into the detector at a constant level optimized for high fidelity demodulation. This limiting effect reduces the impact of any remaining noise bursts, which would affect an AM receiver. As a receiver moves far from the transmitter, the signal amplitude into the detector remains constant, but increasing amounts of the energy are coming from amplifying the noise floor. As the signal weakens, the output to the speaker remains at a constant loudness, as it slowly transitions from a clear signal to the hissing sound of amplified noise.

FM demodulators historically have Foster-Seely discriminators, where the IF signal is sent to two LC filters, tuned for the low and high ends of the 200 kHz bandwidth (Figure 13). Each is followed by an AM detector. When the instantaneous frequency is low, the output of the low end filter increases and the high end filter decreases. For higher frequencies the output of the high end filter increases and the low end decreases. The outputs of the two AM detectors are summed. Proper adjustment of the L-C filters will yield a linear detection characteristic over the 200 kHz bandwidth.

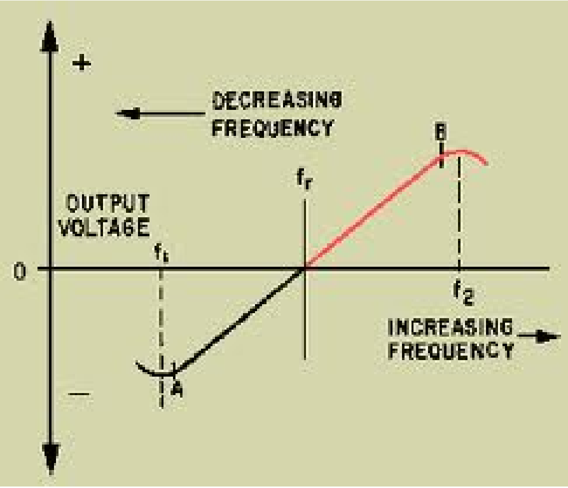

Figure 13 - FM demodulation characteristic

When a FM signal passes through a discriminator, the instantaneous output voltage varies with frequency, but the average point is at zero only when the signal is perfectly tuned in. An automatic frequency control feedback loop (AFC) averages the discriminator output and uses any non-zero average to slowly drive the tuning oscillator in a direction to keep the average output near zero. This system, called Automatic Frequency Control (AFC), is used to pull the receiver in to the exact center frequency for an undistorted output without painstakingly manual fine-tuning.

A modern variation of an FM detector uses a phase lock loop, similar to the synthesizer discussed before, but simpler because there are no frequency dividers. Here, a free-running voltage controlled oscillator (VCO) is brought to a phase detector just as in the synthesizer. However, the other input of the phase detector is the signal coming out of the IF amplifier. Again, the phase detector output is filtered with a filter matched to the instantaneous input frequency swings, and fed back to the VCO to follow the IF output frequency. The VCO follows the instantaneous input frequency as it varies with the modulation, and the VCO control voltage is an accurate representation of the original modulation waveform containing the program.

There has always been tremendous confusion about the noise reduction capability of FM. In the FM band, the atmospheric noise floor from distant thunderstorms is basically nonexistent, so the noise floor drops to the theoretical limit of -114 dBm / MHz due to molecular activity in the receiver hardware. Obviously, the wider a communication channel, the more opportunity for any atmospheric noise to enter, and FM broadcasts are no exception. For that reason, everything else being equal, an FM signal will fade into the noise floor before a (narrowband) AM system will. However, unlike AM, FM is completely immune to noise crashes from local thunderstorms.

4.2 Walkie Talkies

Much portable communication (the ubiquitous handheld radio) is at this frequency band. Such radios use narrow band FM, which uses FM techniques at approximately the same bandwidth of HF band AM signals. Here the differences between AM and FM are reduced. One key factor is that the high density of channelized communications systems in cities and other high population density areas forces narrowband signals.

Handheld radios are of limited range. There is no atmospheric propagation to speak of beyond line of sight, and to maintain battery life transmitter power must be fairly low. Then, the small antennas required to keep them portable do not radiate well; add in all of the surrounding metal and concrete of a big city. Also range is an issue.

Because of the narrowband FM, the receivers have some characteristics closer to AM radios than FM. The FM demodulators are still there, but the IF filters need to be about 15 kHz wide; a little wider than the 10 kHz of AM broadcast, but much narrower than the 200 kHz of FM broadcast. Receivers often are dual conversion with wide filters at 10.7 MHz followed by narrow filters at 455 kHz.

The input stages of such VHF and UHF radios tend to have very small L-C components at these high frequencies. Inductors are often a few turns of wire with a ¼ inch diameter and no core. Some higher-end radios may transition entirely from L-C circuits to resonant transmission lines or resonant cavities, where the dimensions are directly related to the wavelength of the signal. Such devices can produce excellent filter characteristics.

4.2.1 Repeaters

Handheld radios are often supported by permanent stations called repeaters. A repeater will receive on one frequency and instantaneously re-transmit the received modulation onto a slightly different frequency. Repeaters can use two antennas; one to receive and one to transmit, or a single common antenna with a special filter called a duplexer to isolate the transmitted and received signals from each other. Duplexers are usually fairly large, and built very precisely to ensure sufficient isolation between the transmitter and receiver frequencies within the same antenna.

The purpose of this exercise is that each handheld unit can communicate with a fixed unit with a good, high antenna far better than it can with any of the other handheld units, especially in a city environment. With a repeater, it is as if someone on a hill is relaying communication between all the handheld units below.

5.0 Data Communication

Radios are not only used for voice. Various techniques are used to communicate prepared text.

5.1 Radio Teletype (RTTY)

During World War II, a need arose for text transmission via radio; human operators were labor intensive and subject to errors. Out of that need came Radio Teletype (RTTY) which uses frequency shift keying (FSK). Originally only found on the HF band, it has moved up to higher frequencies

Digital data transmission predates radio by 60 years; the Morse code used by telegraph operators since the 1840's is a digital data transmission medium (it can even be considered "texting"). Morse code was designed to be decoded by the human brain (the common letters were assigned simple codes - E being a single "dit" - while the least common letters have 4 symbols). Operators also vary slightly in their sending of code (their "fist" or "swing"), and while the human brain sorts it out well, not so much for automated decoding.

Teletype uses codes where each character has the same number of symbols to enable proper synchronization by a circuit or a computer. The first code was Baudot, where one code represents the shift-lock key to indicate the following characters are capitalized. By that technique, a 5–bit code was able to carry the upper and lower case alphabet, numbers, and some punctuation. Later, as transmission speeds increased and allowed for a more flexible code, the 7-bit ASCII code became the standard.

Frequency shift keying is a series of ones and zeros, just like any modern digital transmission medium. In this case, zeros are indicated by a slight decrease in signal frequency and ones are indicated by a slight increase in frequency. Synchronizing bits are used to notify the receiver when characters begin and end, but the transmission must contain roughly similar numbers of 1's and 0's to make sure the average frequency remains on-channel.

The standard shift at first was 170 Hz, although that number has changed as new codes have been developed in recent years. While new codes and modulation techniques have been introduced, the overriding concern for any digital transmission medium is to maintain the data rate such that the signal bandwidth will stay within the assigned channel width. For HF, that is the same width as a communications-level voice transmission, where tuning in an RTTY signal yields a unique warbling sound.

5.2 Telemetry

Other forms of data communication include telemetry, historically used for monitoring rocket launches, missiles and other unmanned vehicles. In recent years, military drones have state of the art communication links, but such links still harken back to the basics.

Classic telemetry receivers are used to monitor such things as rocket launches. One interesting use is for range safety. We have all seen a failed rocket launch on the news with the announcer sadly intoning "Range safety had to destroy the rocket." Such a data link involves a telemetry receiver on the rocket operating at roughly 450 MHz, using AM modulation of typically three audio tones. In an interesting application, the destruct message requires a very specific sequence of the tones. Normally, readiness is verified by sending different tones, but being incredibly careful to stay far away from the destruct sequence. The receiver outputs which tones are received to a separate telemetry transmitter tasked with sending many monitored parameters to the ground. In that manner reception of the tones is constantly verified to be assured that the destruct system will always work on command.

5.3 Digital modulation

More modern telemetry involves digital modulation with very high data rates, usually with computer-generated packets of data. It is now common to digitize an analog signal, store it, bundle it onto a packet, and send it out all at once. Successive packets must line up with the previous and following packets perfectly for the system to work. Obviously it can be done well, because virtually all telephone calls are now handled that way, even when totally ground-based.

There are many different modulation schemes. The only real criterion is that all stations on a network use the same protocol. The original frequency shift keying approach of teletype is actually used in one new military datalink, where the (large) frequency shift keying changes the frequency the same amount as the (high) keying rate (Minimum Shift Keying or MSK).

5.3.1 Phase Modulation

More typically, data communication uses some form of phase modulation. There are many types of phase modulation, although the basic principles are the same.

5.3.1.1 Differential Phase Shift Keying (DPSK)

A basic form of phase modulation is DPSK (differential phase shift keying), where what appears to be a continuous CW carrier is sent. Upon closer examination, there are periodic inversions (180 degree phase shift). The inversion represents transmission of a logic 1, while a logic 0 is indicated by no change at all. Since the data rate is well defined, every time the inversion of a (logic 1) is received, the positions where the next few candidate bits are allowed to change are predicted accurately. If there is an inversion at the expected time, it represents a 1; no change is a 0.

Since the carrier is otherwise unmodulated, converting the signal to digital is not difficult. The classic approach is to digitize the raw analog signal by setting a threshold at the middle of the amplitude, then forcing voltages above that level to logic 1, and those below to logic zero. This so-called “hard-limiting” can be interpreted as a 1-bit A/D conversion. If the signal is a reasonably low IF frequency, it can be decoded by passing it through a digital circuit to delay it by precisely the bit rate. The input and output of the delay are compared to output ones when there has been a phase inversion, and zeros when there hasn’t.

This technique works quite well in noisy communication channels, as after a few bits are decoded, the start and end times of the upcoming bits are well known. This information can be used to filter out noise and other transient variations to improve decoding even in a noisy environment. One approach is the integrate-sample-dump, where the signal is averaged (integrated) between samples, so that much of the noise and other variations will cancel out. Once the time comes to decide on that bit, the average is sampled and the result sent out. Then the integrator is reset for a fresh start on the next bit. (Note that this technique is usually called “integrate and dump”, but shouldn’t we do something with the result before we throw it away?)

At the cost of needing a better signal-to-noise ratio, differential quadrature phase shift keying (DQPSK) allows for phase shifts in multiples of 90 degrees to provide two bits of data.

5.3.1.2 Using a Phase lock loop for phase detection

The phase lock loop described earlier to generate precise, yet selectable, local oscillator frequencies, and as a FM demodulator, can also be adapted to serve as a phase detector (no surprise; the name is a big hint). Much like the FM demodulator, a voltage controlled oscillator operates over a narrow frequency range around the receiver IF frequency. The phase detector compares the VCO with the IF output and adjusts the VCO to match not only the frequency, but also the phase of the input signal. If the phase of the input signal changes appreciably faster than the feedback loop can follow, the slow loop will cause the VCO to settle at the long-term average frequency of the input. Meanwhile, the instantaneous output of the phase detector follows the phase shifting of the input.

5.3.1.3. Forward Error Correction

When digital messages are sent at a fairly high data rate, one can afford the luxury of reducing the data rate slightly to add forward error correction, which is a data redundancy that is added to detect – and even correct – some errors in decoded data.

An obvious example, although just as obviously inefficient, is to send each message three times. For any data bit in the message, should one of the three repetitions differ from the other two, the majority will rule. Obviously, more efficient methods have been developed to provide error correction over noisy or otherwise corrupted communication channels. Even then, such techniques can only correct for a limited amount of degradation before the channel becomes unusable. In systems where there is full-time, two-way communication, usually protocols are set up such that when a data packet is determined to be corrupted, a request to send it again is sent back to the originator.

5.3.1.4. Spread Spectrum

Spectrum spreading is a technique where a transmitter adds high frequency components to a signal to intentionally make its spectrum much wider. If the high frequency components have a quasi-random characteristic, and the spectrum is spread widely enough, the signal will blend in with the noise floor and become harder for the uninformed to even detect its presence. Encrypting the pseudo-noise spreading signal makes it even more secure, but non-encrypted spectrum spreading is often used in commercial applications such as GPS to allow the incredibly weak GPS signals wafting down from space satellites to be accurately decoded by that little box on your windshield, and without the need for any large antenna arrays. (Now, that's action at a distance!)

5.3.1.5 Phase Shift Keying (PSK)

Many modern digital communication formats such as cellular phones, terrestrial data links and especially satellite communications need extremely high data rates, and use quite high frequencies and elaborate modulation schemes.

Data links using phase shift keying can handle single bits (Binary PSK or BPSK), two bits (Quadrature PSK or QPSK), or even three bits (8-PSK). This requires a more complex receiver with In-phase and Quadrature (I, Q) outputs, but the error rate improves.

An I-Q detector is usually a variation of a phase lock loop detector, with one oscillator providing two outputs, 90 degrees out of phase, to two phase detectors. The phase detector whose feedback closes the loop becomes the "I" channel, and the one running open-loop, 90 degrees out, is the Q-channel.

In phase modulation, the signal includes magnitude and phase. If the last local oscillator (analogous to the BFO used for demodulating CW) provides two outputs 90 degrees out of phase (in quadrature) to two mixers, then the zero degree mixer provides both feedback to close the loop and the In-phase (I) output, while the 90 degree mixer runs open-loop and provides the quadrature (Q) output.

The two demodulated ("baseband") channels are digitized and processed to recreate the phase information. One way of looking at it is to show the phase as a "constellation diagram", which consists of a circle (0-360 degrees of phase), with constant radius (amplitude). The horizontal axis shows the I-channel and the vertical output shows the Q-channel. Figure 14 shows a common constellation diagram for 8-PSK.

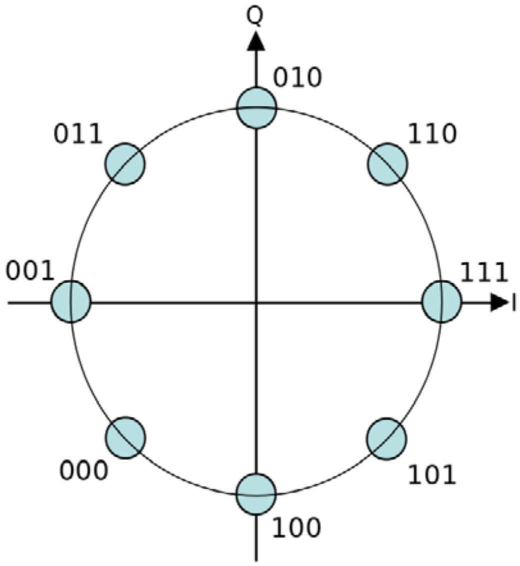

Figure 14 - 8-PSK constellation diagram

5.3.1.6. 16-QAM

Further increases in resolution of digital data streams come from using both amplitude and phase modulation (at a cost in susceptibility to corruption). In 16-QAM, the constellation diagram expands on the 8-PSK diagram by adding amplitude variations to the phase (8-PSK uses one amplitude and eight phases while 16 QAM uses four amplitudes and four phases). The constellation diagram expands to that shown in Figure 15.

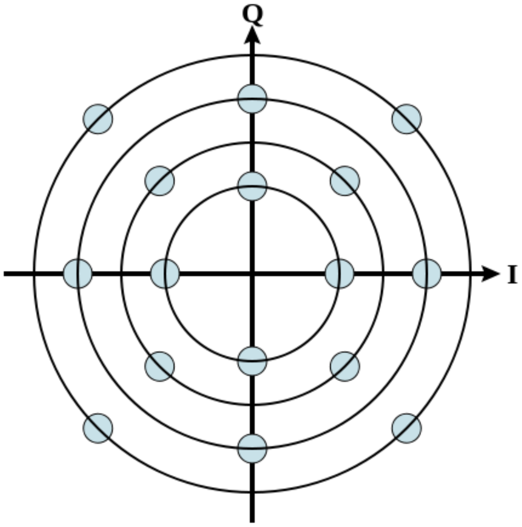

Figure 15 - 16-QAM constellation diagram.

5.3.1.7 Higher QAM modulations
In recent years the number of amplitude and phase steps used in Quadrature Amplitude Modulation has continued to grow exponentially, with increases to 32 QAM, 64 QAM, and even 256-QAM. With 16 amplitude steps, and 16 phase steps, the constellation diagram appears quite busy.

6.0 Microwave Receivers

Usually communication systems at frequencies above 1 GHz or so are highly directional. The efficiency of an antenna to pick signals out of the Aether depends on its size. At frequencies above 1 GHz simple antennas with broad directional coverage become quite small and do not intercept much passing stuff. Combining multiple antennas improves performance, typically in a highly directional manner. Such highly directional antennas are extremely useful for applications such as tower-to-tower microwave links, radar, etc.

Receivers at these higher frequencies do not differ appreciably from those already discussed. Here L-C components become so small that they are unwieldy, while resonant cavities and other devices become practical and replace the L-C circuits, often with sharper bandwidths due to the lower losses.

Sometimes a microwave receiver is split, with the “front end” located right at the antenna to minimize losses in cables. Alternatively, microwaves often use waveguides to provide low losses, although they are rigid. Placing the RF amplifier and the first conversion stage right at the antenna allows for a simpler way to bring the first IF to the remainder of the receiver.