HF Radio Communications

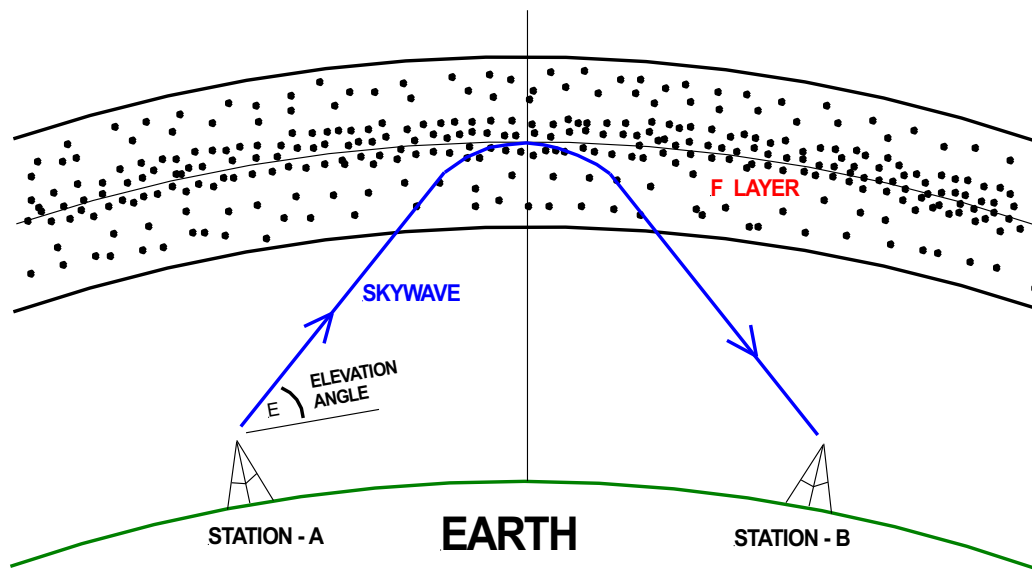## Chapter 1

SKYWAVE

F LAYER

ELEVATION ANGLE

E

STATION - A   EARTH   STATION - B

1   Introduction

High Frequency (HF) radio communications is important! HF often provides dependable service during disasters when all other forms of communications, including cell phone, internet, and even satellite communications, have failed. In remote areas, particularly mountainous regions, HF is usually the only form of communications available. HF communications is unique in that it is primarily a stand-alone system. The only required infrastructure being Earth's ionosphere which is free, available everywhere, and useful most of the time. In contrast, the other forms of electronic communications require large, complex, extremely expensive infrastructures costing hundreds of billions if not trillions of dollars.

Operating in a frequency range from 3 to 30 MHz, HF radio signals propagate through the ionosphere in Earth's upper atmosphere covering distances from a few miles to thousands of miles. This is known as skywave propagation. There are also two other forms of HF propagation useful for very short local communications. These are line-of-sight and ground wave propagation. Line-of-sight propagation covers distances up to 20 miles or so while ground wave propagation has a range of up to 50 miles depending on the operating frequency and ground conditions.

There are problems that must be delt with to achieve successful HF radio communications:

- The characteristics of the ionosphere are continuously changing throughout the day, seasonally, and with the 11-year solar cycle requiring operating frequencies to change accordingly.

- Solar x-ray radiation, high energy particle, and solar wind storms can disrupt HF communications

- Atmospheric, galactic, and man-made noise may cause radio signals to be lost in the background noise

- The large number of possible propagation paths between transmitting and receiving stations can result in multipath interference problems at the receiving site

- Various forms of signal fading can make HF radio communications difficult as received signals repeatedly peak, fade below the noise level, reappear, and fade again.

An in-depth understanding of HF radio communication principles and operation allow these problems to be successfully handled, making possible reliable HF communications

- Anywhere,

- At any time,

- Under most conditions.

Finely, thousands of amateur radio operators around the world enjoy HF communications as an exciting hobby.

Radio signals are electromagnetic waves predicted by James Clerk Maxwell in 1865. Radio waves and light waves are the same thing differing only in wavelength. Radio waves are at the bottom of the electromagnetic spectrum with wavelengths of more than 100 meters as illustrated in Figure 1. Visible light wavelengths range from 700 down to 400 nanometers (nm), while Gamma rays are at the top of the spectrum with wavelengths on the order of 0.0001 nm.
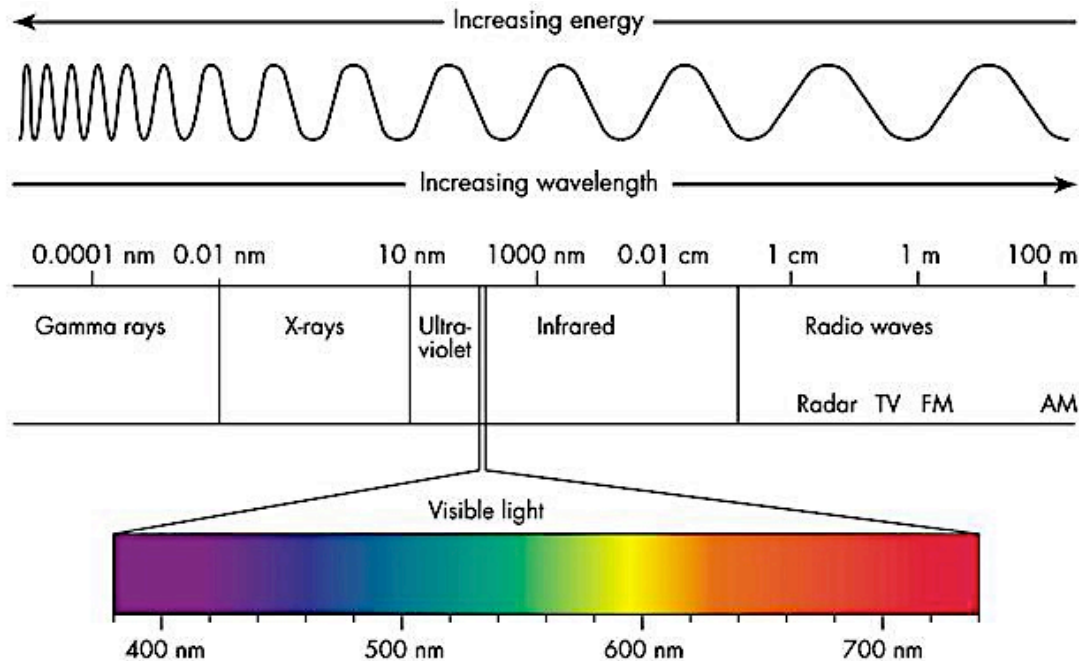

Figure 1 Electromagnetic Spectrum   (source: Cyberphysics)

Radio signals can travel through empty space from one place to another, the same as light waves.

Hertz proved the existence of the radio waves predicted by Maxwell in a series of carefully conducted experiments in 1888.

Marconi, only 20 years old at the time, transformed the experiments of Hertz into a commercially viable product ushering in the wireless telegraph vital for ship to shore and ship to ship communications.

In late 1924 British scientist Edward Victor Appleton demonstrated that radio signals propagated though an ionized region of the atmosphere approximately 80 – 90 km above the Earth. This region, later called the ionosphere, bends radio signals back to Earth making possible both short and long-distance radio communications.

Around the same time amateur radio operators discovered the long-distance propagation capabilities of HF radio. Commercial radio companies, including the British Broadcasting Company (the BBC), flocked to the HF frequency bands providing worldwide communication and broadcast service.

During World War II HF radio quickly became the primary means of communications in Europe as telephone and telegraph lines were cut. HF, both encrypted and non-encrypted, was the only means of communications with hundreds of ship convoys, long distance aircraft, and thousands of islands throughout the vast Pacific theater. HF communications truly came of age during WWII with explosive expansion in HF technology.

1.1   James Clerk Maxwell (1831 – 1879)

For centuries magnetism and electricity were believed to be completely different phenomena. In the early to mid 1800s experiments performed by Oersted, Faraday, Henry and others seemed to indicate that magnetism and electricity were somehow related. In 1865 Scottish physicist and mathematician James Clerk Maxwell, Figure 2, published a set of 20 equations that unified electricity and magnetism into a single theory of electromagnetics. Later Oliver Heaviside, using vector calculus, simplified Maxwell's equations into the set of 4 differential equations that we use today.
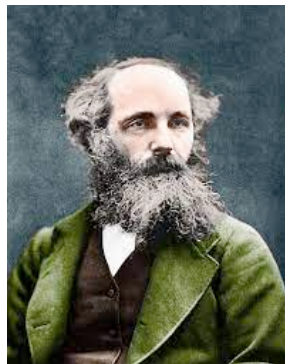
James Clerk Maxwell   Oliver Heaviside

Figure 2 Maxwell and Heaviside (source: Wikipedia)

Maxwell is regarded as one of the greatest theoretical physicists of the 19th century. Maxwell died young at the age of 48. However, in this short life, he formulated the unifying theory of electromagnetics, developed the kinetic theory of gases, made significant contributions to our understanding of color vision and studied the nature of Saturn's rings.

In developing his set of equations, Maxwell integrated together the following laws of electricity, magnetism, and induction originally developed by Gauss, Faraday, and Ampere.

Gauss's Law:

$$\oint \vec{E} \cdot \vec{dA} = \frac{Q}{\varepsilon}$$

Gauss's Law of Magnetism:

$$\oint \vec{B} \cdot \vec{dA} = 0$$

Faraday's Law of Induction:

$$\oint \vec{E} \cdot \vec{dl} = -\frac{d\phi_B}{dt}$$

Ampere and Maxwell's Law

$$\oint \vec{B} \cdot \vec{dl} = \mu_0 I + \varepsilon_0 \mu_0 \frac{\partial \phi_E}{dt}$$

Maxwell's equations showed that it was possible for time varying electric and magnetic fields to be self-sustaining, each inducing the other. That is, a time varying magnetic field could induce a time varying electric field which in turn induced the magnetic field. Based on this result, Maxwell predicted the existence of electromagnetic waves, each wave consisting of an alternating electric wave perpendicular to an inseparable magnetic wave, and both orthogonal to the direction of wave travel as illustrated in Figure 3. Since each wave induced the other, the electromagnetic wave itself could theoretically propagate forever in its direction of travel through empty space.
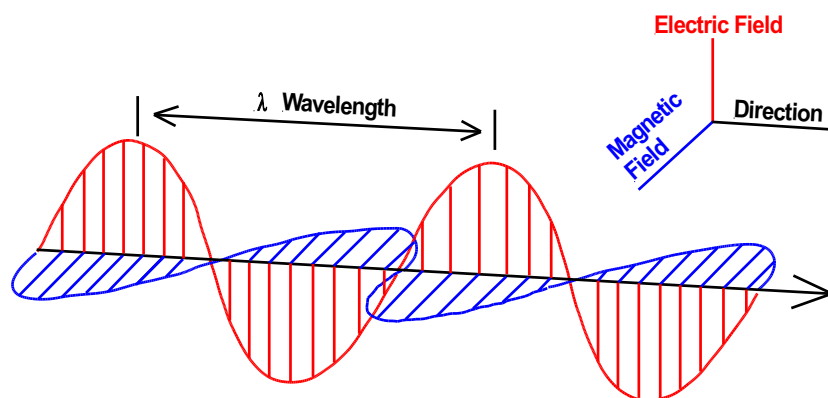

Figure 3 Electromagnetic Wave (source: author)

Maxwell noticed that the inverse square root of electric permittivity in free space, ε_0, and free space magnetic permeability, μ_0, had the units of velocity. Maxwell concluded that the velocity, v, of his proposed electromagnetic wave must be

$$v = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$$

Coulomb and Ampere had previously determined the values of ε_0 and μ_0 through careful static electric and magnetic measurements. Using their measurements Maxwell calculated the speed of an electromagnetic wave to be 284,000,000 meters per second, a value very close to the speed of light known at the time. Maxwell concluded that light itself must be an electromagnetic wave.

1.2   Heinrich Hertz

Figure 4 Heinrich Hertz (source: Wikipedia)

20 years later in 1888 physicist Heinrich Hertz tested Maxwell's equations to see if electromagnetic waves really existed. To verify Maxwell's equations Hertz had to demonstrate three things:

1. A rapidly changing electrical current, such as that produced by an electric spark, would generate an electromagnetic wave of the type predicted by Maxwell,

2. The electromagnetic wave would travel through space from one place to another, and

3. That it would do so at a finite speed near the speed of light.

Hertz used the apparatus shown in Figure 5 to perform these experiments. The apparatus worked as follows. A relay device, called an interrupter, placed in series with the battery "chopped" the electrical current flowing from the battery through the primary winding of the induction coil into electrical pulses. The pulses produced voltage spikes on the order of several thousand volts in the induction coil secondary winding. A capacitor consisting of two metal sheets was connected across the induction coil secondary as shown in Figure 5. The capacitor was charged to a very high voltage by the secondary winding voltage spikes. When charged to a high enough voltage, air molecules in the gap between two closely spaced spherical balls were ionized. The ionization allowed an electrical current to flow across the gap from one ball to the other producing a high voltage spark.

The voltage spark discharged the capacitor. The process of repeatedly charging and discharging the capacitor continued as long as the switch in the battery circuit was closed. The capacitance of the two metal sheets plus the inductance inherent in the secondary circuit wiring formed a series resonant circuit which in the experiments conducted by Hertz was around 150 to 200 MHz. This circuit satisfied the first of the three conditions needed to verify Maxwell's equations.

The second condition required verification that the electromagnetic wave presumably produced by the spark gap circuit did indeed travel across Hertz's laboratory. His laboratory was a large lecture hall. To verify this condition Hertz constructed a wire loop with a very small air gap as shown in Figure 5. Hertz held the loop at the opposite end of the lecture hall from his radiating spark gap transmitter. When he did so, small sparks developed across the air gap in the receiving loop verifying that electromagnetic waves were in fact traveling across the laboratory.
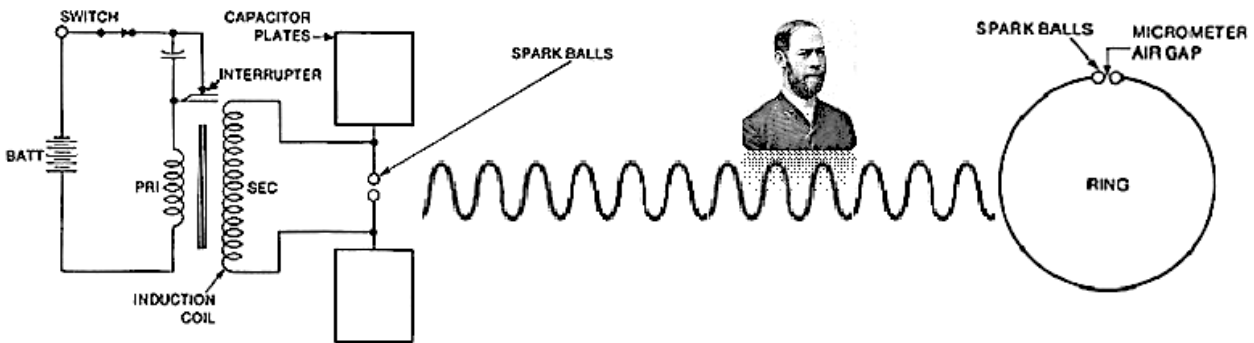

Figure 5   Experimental apparatus used by Hertz. (source: Wikipedia)

The third step, verifying that the electromagnetic waves were traveling at the speed of light was more difficult to prove. To do this, Hertz placed a large metal sheet on the wall at the far end of the lecture hall from his transmitter. The hope was that electromagnetic waves reflected from the metal sheet would interfere with the waves radiating from the transmitter forming standing wave patterns exactly the same as those obtained in reflecting light. To detect the standing waves, Hertz walked toward his transmitter holding his receiving loop and observing the strength of the sparks occurring across the receiving loop's air gap. As he did so he observed strong sparks across the air gap at certain distances from his transmitter, indicating standing wave voltage maximums. At other locations there were hardly any sparks at all signifying standing wave minimums. The distance from one peak to the next represented the wavelength λ of the standing wave. Using this result, and estimating (guessing at) his transmitting frequency f, Hertz was able to calculate the velocity v of his electromagnetic wave according to

$$v = \lambda \cdot f$$

which he found to be nearly equal to the speed of light c. While this experiment was admittedly crude, it verified that electromagnetic waves did travel through space at a finite speed which was on the order of the speed of light.

Other scientists at the time were also experimenting with Maxwell's equations at lower frequencies without much luck. What caused Hertz to succeed was that he was operating, intentionally or otherwise, at relatively high frequencies permitting standing waves with two or more peaks to develop within the confines of his laboratory.

Hertz performed his experiments, including development of his test apparatus, over a considerable period of time. It was not a quick couple day project! In the end Hertz verified that Maxwell's equations predicting the existence of propagating electromagnetic waves was indeed correct.

Hertz was a physicist focused on the theoretical aspects of Maxwell's equations. It is unlikely that the commercial ramifications of his electromagnetic experiments ever occurred to him. With his experiments complete, he went on to other things. Hertz died in 1894 of blood poisoning at the early age of 36.

1.3   Guglielmo Marconi

Figure 6 Guglielmo Marconi (source: Wikipedia)

In 1894, when he was 20 years old, Marconi read an obituary on Heinrich Hertz describing the work that Hertz had done. Marconi realized the commercial ramifications of the electromagnetic experiments performed by Hertz and began repeating the experiments at his father's estate (Villa Griffone) near Bologna, Italy. Like Hertz, Marconi initially experimented at approximately 2 meters. At these high frequencies he was unknowingly limiting himself to line-of-sight distances.

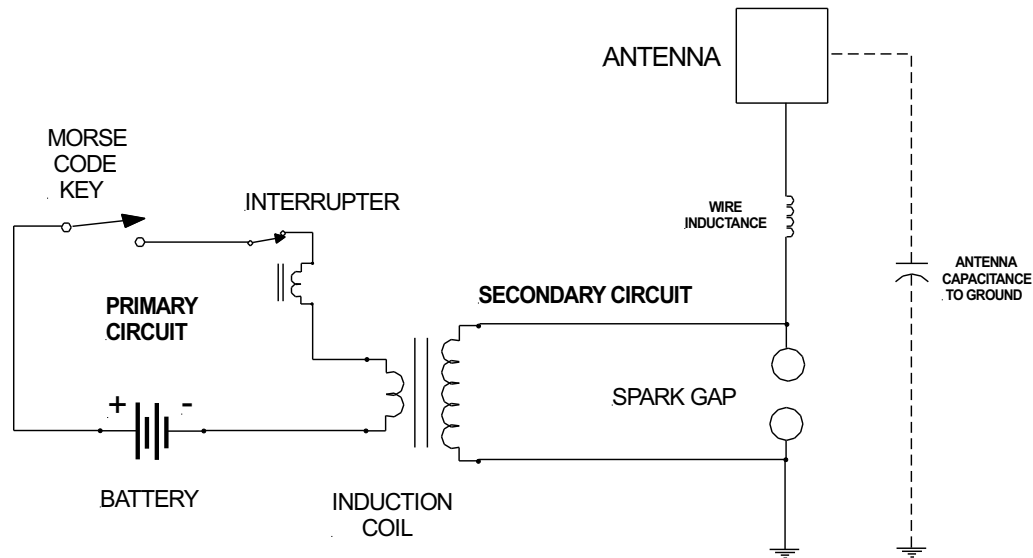Marconi's spark gap transmitter (Figure 7) was pretty much the same as that used by Hertz.

Figure 7 Marconi's first spark gap transmitter   (source: author & Wikipedia)

While his spark gap transmitter was similar to that used by Hertz, Marconi's spark gap receiver, shown in Figure 8, was vastly different than the simple receiving loop used by Hertz.

A practical receiver required a detector that could activate a relay when an electromagnetic wave was detected. Such a device, called a coherer, consisted of a tube containing metal filing filling the gap between two electrodes as illustrated in Figure 9. In the presence of an electromagnetic wave the metal filings became fused together allowing an electrical current to flow through the coherer activating a relay as shown in Figure 8. The relay in turn activated a telegraph sounder, a larger relay, causing the receiving circuit to sound exactly like the clickity clack of a telegraph system when receiving morse code. In Figure 8 the device marked A B C is the coherer, R is the relay, S the sounder, B_1 and B_2 batteries, and L circuit inductance. The glass tube in the picture on the right in Figure 8 is the coherer while the two tan colored cylinders are the relay R.
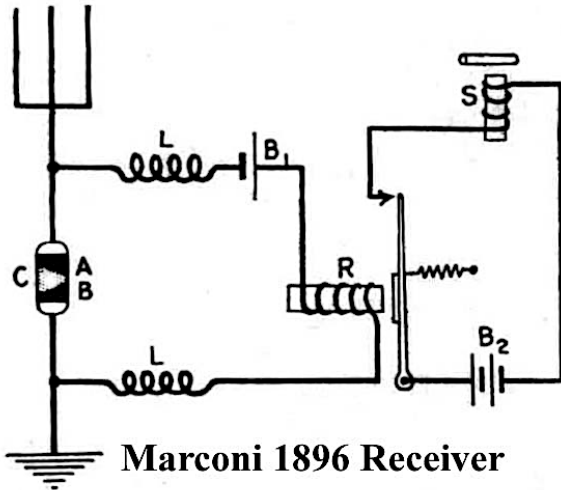
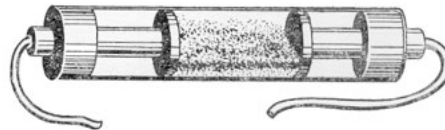Figure 8 Marconi's 1896 spark gap receiver   (source: Wikipedia)

Figure 9 A coherer   (source: electronics.stackexchange.com)

In 1890 Edouard Branly had discovered that the resistance measured across the ends of a tube filled with metal filings was normally very high. But, in the presence of an electromagnetic wave the metal filings became fused together resulting in a very low resistance across the tube. The resistance remained low after the electromagnetic wave disappeared. Tapping the tube separated the metal filings returning the tube to its original high resistance state. The coherer was in affect a switch that was turned on by the presence of an electromagnetic wave and turned off by some sort of mechanical tapping arrangement. Branly was interested in why this occurred. He had no interest in wireless communications. In 1894 Oliver Lodge expanded on Branly's work developing a superior cohere which Lodge used in experimenting with the similarities between light and electromagnetic waves. It was Lodge who gave the device its name.

The coherer that Marconi used in his receiver was an improved version of Lodge's cohere. Marconi's improvement consisted of enclosing the metal filings in a tube partially evacuated of air. For a number of years, Marconi's cohere was the best detector available.

What drove Marconi personally, and in obtaining money from his father and others for his experiments, was transmitting long distances. At first a few feet, then progressively further. In 1894

Marconi transmitted a distance of 2 miles near his home in Bologna, Italy. In March 1899 he transmitted a signal across the English Channel from Dover, England to Boulogne, France. His greatest achievement came on December 12, 1901 when he and his associate transmitted a signal more than 2,000 miles across the Atlantic from Poldhu in Cornwall, England, to St. John's, Newfoundland, Canada.

To Marconi, transmitting longer distances meant using larger antennas. In 1896, as the result of "cut and try" experimentation, Marconi built a vertical antenna by connecting the upper part of the spark gap to a long vertical wire, and the other part to ground, as illustrated in Figure 10. In this figure the transmitting station is shown on the left and the receiving station on the right. Using a vertical antenna considerably increased transmission distance.

Transmitting Station   Receiving Station
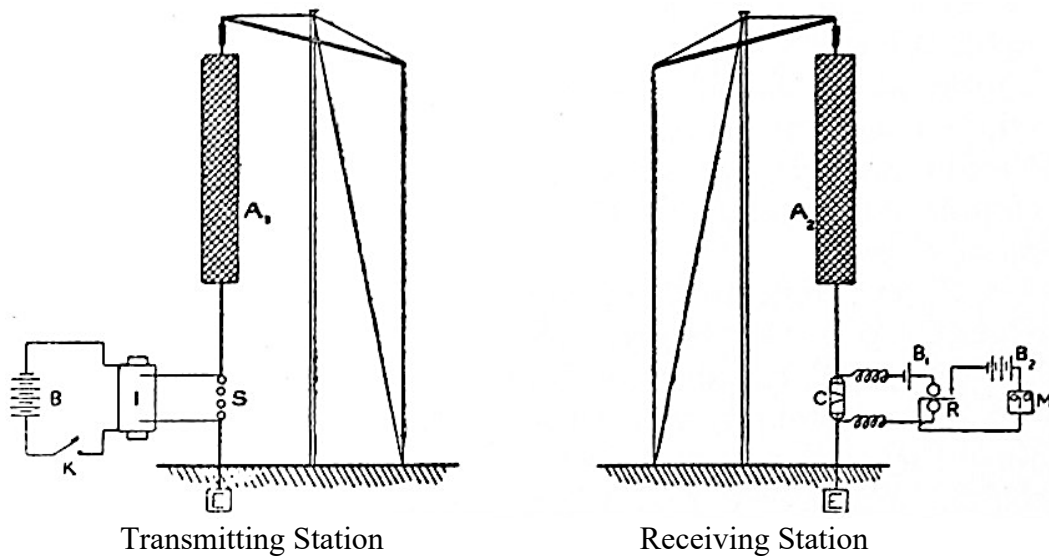
Figure 10 Marconi's vertical antenna (source: Attken)

The rectangular cross hatched object marked A_1 in Figure 10 is a metal sheet forming the upper plate of a capacitor. The lower plate being the ground itself. As illustrated in Figure 5, two relatively large capacitor plates were an integral part of early spark gap transmitters. The object marked A_2 at the receiving station performed the same function. These plates eventually disappeared as the technology evolved. Today's ground mounted vertical antennas date back to Marconi's 1st vertical in 1896.

Increasing the length of the vertical wire lowered Marconi's operating frequency according to the equation

$$f = \frac{c}{\lambda} = \frac{c}{4h}$$

where

c = the speed of light

$\lambda = wavelength$

h = the height of Marconi's vertical antennas which were roughly $1/4 \ \lambda$ tall ($1 \ \lambda = 4h$).

By using lower frequencies Marconi achieved longer ground wave propagation distances, particularly over salt water. Eventually Marconi wireless systems operated at frequencies from 1 MHz down to 300 KHz requiring very large antennas. The ionosphere at the time was of course unknown.

At his father's insistence, Marconi offered his wireless system to the Italian Navy, but they were not interested. Marconi's father was not only wealthy, but his English born mother was part of a well to do well connected family back in England. In the spring of 1896 Marconi and his mother traveled to England. With access to the right people, Marconi quickly began discussions with the English Post Office. They were very interested in Marconi's wireless. By law the English Post Office was responsible for all electrical communications in England meaning that all telegraph business in England was under its control. To William Preece (chief engineer at the Post Office) the fledgling wireless technology would also become part of his jurisdiction. Others including Oliver Lodge were experimenting with wireless. But to Preece, young Marconi seemed to be a very promising prospect for developing a workable wireless communications system. Wishing to control the evolving wireless business, Preece threw his support and the prestige of the Post Office behind Marconi providing Marconi with considerable technical and political support.

Marconi's cousin, a very capable business man, strongly suggested that Marconi start a wireless communications company which Marconi did. It was a wholly own family business financed by his extended English family. Because of his well to do family, Marconi had opportunities that no one else involved in wireless had. Marconi did not have to show immediate profits. His family was willing to finance him with the expectation that sometime in perhaps the distant future their investment would pay off, as it eventually did.

Marconi could not compete with the wired telegraph. The extensive world-wide telegraph networks, depicted in Figure11, were owned and operated by large well-established companies. The wired telegraph networks were far more dependable than the wireless. In addition, morse code speeds were very fast over wired telegraph compared to the slow speed wireless.

Throughout most of history, all knowledge of a ship was lost from the time it left port until, hopefully, it arrived at its destination. Once at sea, its location and its situation were completely unknown (Figure 12). Ship-to-Shore and Ship-to-Ship communications offered Marconi a golden opportunity which the telegraph business could not compete with. Ship owners were extremely interested in the wireless. But they did not want to buy hardware from Marconi. They wanted Marconi to provide the complete system including shore-based radio stations plus Marconi build, installed, and operated radio equipment aboard ship.

Figure 11. World wide telegraph networks in 1891   (source: Wikipedia)

Figure 12 Situation of a ship at sea was unknown through most of history (source: transpress nz)

Marconi's new company did exactly that. For a while in the early 1900s Marconi achieved nearly a complete monopoly in maritime communications. Many ships were equipped with wireless communications by 1908. However, radio equipment on passenger ships were more for the convenience of wealthy passengers wishing to conduct business while at sea than for safety. That all changed with the sinking of the Titanic. Old time maritime morse code radio operators described wireless nets between ships as one big "party line". A typical shipboard radio room in shown in Figure 13.

Figure 13 Radio room on the S.S. Chindwin (source: Brown's Signaling)

1.4　Antennas

Vertical antennas were the best long-distance antennas. While they were extensively used on shore, they were not practical for ship board operation because of their extensive height. Typical early vertical antennas are shown in Figure 14. Notice the height of Marconi's first 1901 Cornwall Antenna intended for his transatlantic experiment. The antenna is huge because of its low operating frequency. This antenna blew down in a heavy wind storm and was replaced by a sturdier fan vertical like that shown below.
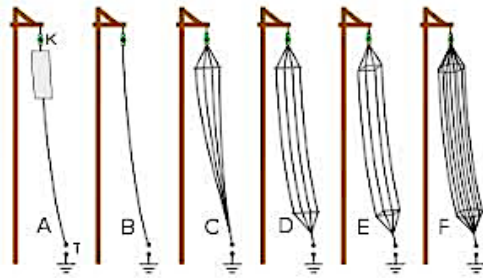
Various vertical wire antennas
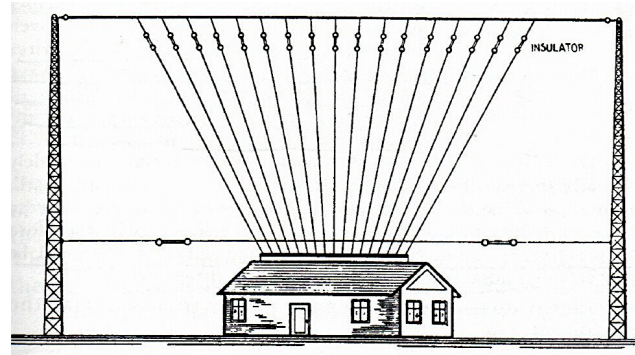
A vertical fan antenna
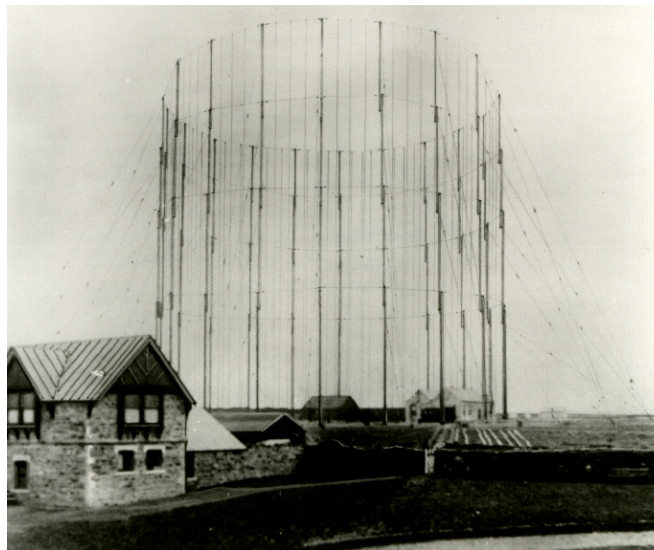
Marconi 1st Cornwall Antenna 1901

Figure 14   Early vertical antennas   (source: Wikipedia)

Inverted L and T antennas were the most popular for ship board operation. Examples of early inverted L antennas are shown in Figure 15. Some land based inverted L antennas, such as the lower antenna in Figure 15, were as much as a mile long. Figure 16 depicts some early T antennas. The vertical, inverted L and T antennas were all $1/4$ λ radiating elements requiring an extensive ground to completed the antenna. Aboard ship the steel hull provided the needed ground. On shore elaborate radial systems buried in the ground were installed. Nearly all of the early antennas used multiple parallel wire assuming that they would radiate more power and provide better capacitance between the antenna and ground than single wire antennas. Although that assumption is questionable, multiple parallel wires are often used on 80 meters to increase antenna bandwidth. One wavelength at the frequencies being used (1 MHz down to 300 KHz) ranged from 900 to over 3,000 feet. Inverted L and T antennas, even very tall ones, were only a small fraction of a wavelength ($0.06\ to\ 0.2\ \lambda$) above ground. Consequently, these antennas were all Near Vertical Incident Skywave (NVIS) short range antennas with very high radiation angles as illustrated in

Figure 17. Inverted L and T antennas are still used today by amateur radio operators on 160 and 80 meters.
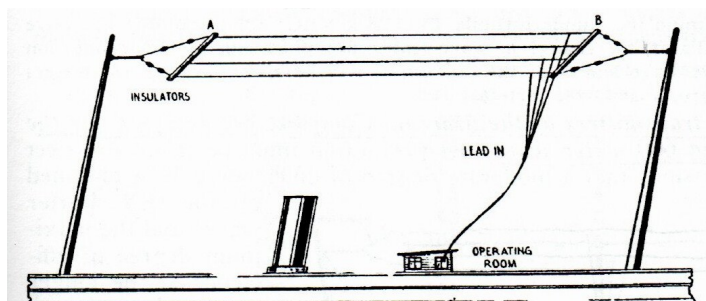
Ship board inverted L antenna   Land based inverted L antenna
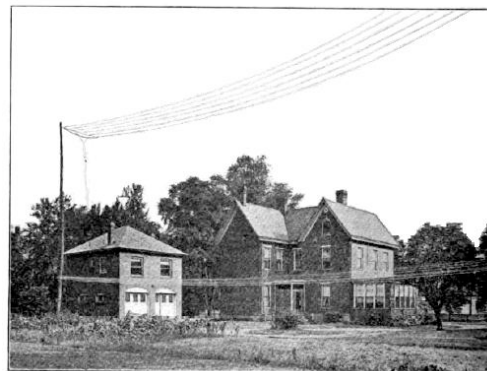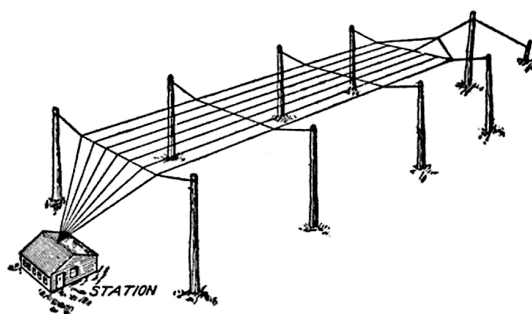

Figure 15   Early inverted L antennas   (source: Bucher)
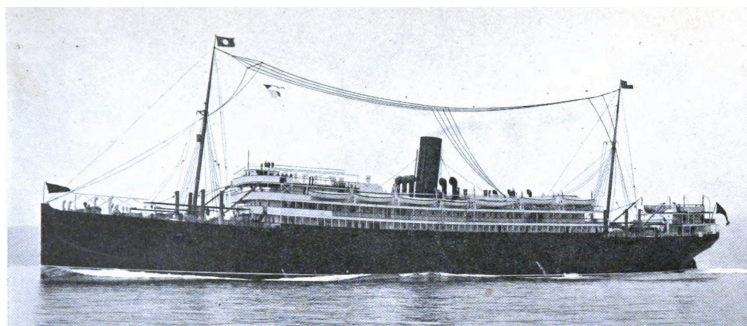
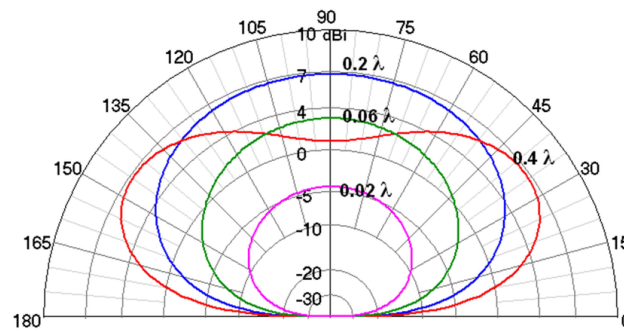

Figure 16 T antennas   (source: Wikipedia)

Figure 17   NVIS antenna vertical radiation patterns   (source: Researchgate)

1.5   Long Distance Low Frequency Communications

Spark gap technology improved from 1900 through 1915. Other electro-mechanical transmitters also evolved. Arc transmitters were the most important alternate technology. Unlike the spark gap, an arc transmitter, illustrated in Figure 18, transmitted a continuous wave (CW) signal. Spark gap transmitters, with their highly damped output signals (Figure 19) were excellent for wireless telegraph, but they could not transmit voice. Arc transmitters with their continuous wave outputs could, and by 1910, were transmitting voice as well as wireless telegraph. But there was no market for wireless telephone until the 1920s when commercial broadcast radio stations emerged.

By 1915 arc transmitters were the most powerful transmitters in the world. Built under contract to the Navy, the Federal Company in Palo Alto, California built arc transmitters (Figure 20) ranging in power from 100 to 1,000 KW and operating, like all commercial transmitters of the day, at 1,000 KHz down to 300 KHz. Arlington, Virginia (Figure 21) was the hub of the Navy's network. Arlington could contact Panama and San Francisco at night and sometimes during the day. Hawaii could be reached by San Francisco and sometimes directly from Arlington. Guam and Hawaii talked regularly, all before amateur radio operators discovered short wave radio.
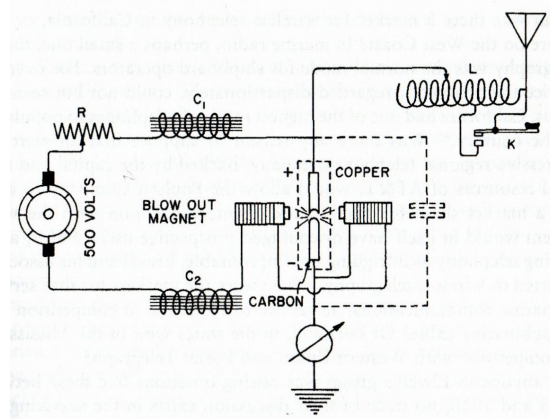

Figure 18   Arc Transmitter   (source: Bucher)

Highly damped spark gap transmitter signal   Continuous wave
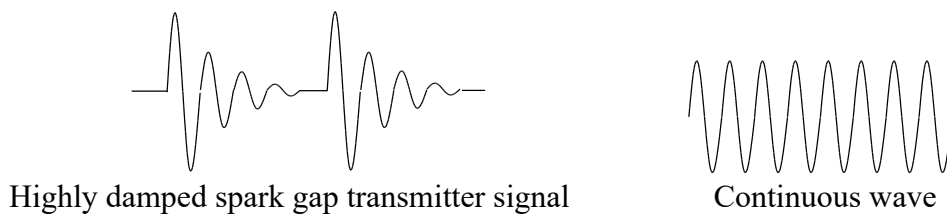
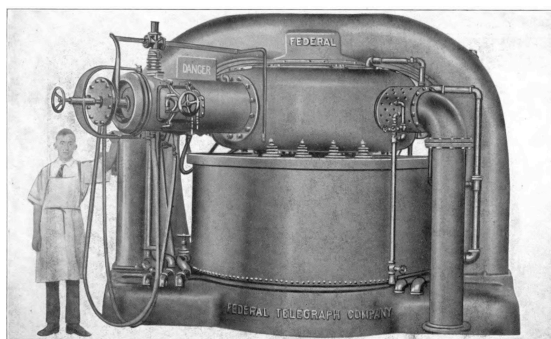Figure 19 Highly damped and continuous waves   (source: author)

Figure 20   Federal 100 KW Arc transmitter weighing around 60 tons   (source: SlideShare)
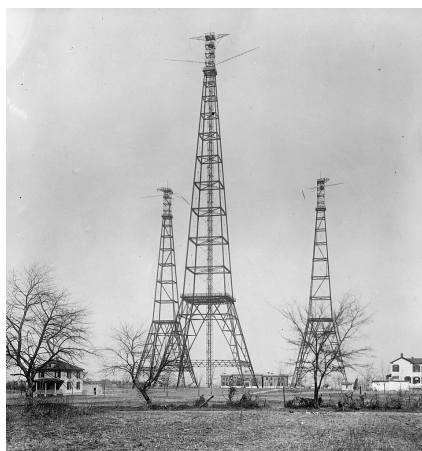
Figure 21 Navy's Arlington, Va. radio station antennas   (source: Pinterest)

1.6   Better Receivers

The coherer was a miserable detector, but for a long time no one could come up with a better device. Eventually it was discovered that a small piece of carborundum crystal could be used in place of a coherer to sense the presence of spark gap signals. By 1906 the coherer had been universally replaced by carborundum crystal detectors which were simple, low cost, and very effective. Crystal detectors allowed anyone with even minimal mechanical skills to build crystal radio sets dramatically increasing the number of amateur radio operators. Figure 22 shows the schematic of a typical crystal radio receiver.
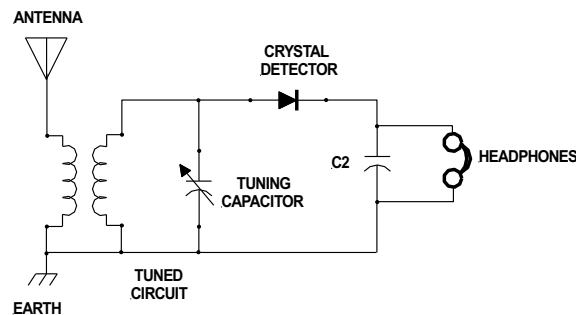

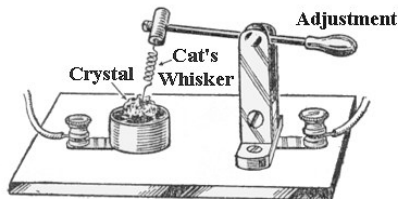Figure 22 Crystal Radio Receiver   (source: author)

Figure 23 Crystal diode   (source: Wikipedia)

A wire (cat's whisker) in contact with a piece of carborundum crystal (Figure 23) formed a diode which in series with the capacitor C2 (Figure 22) demodulated spark gap wireless signals allowing them to be heard in the headphones. The operator probed around the surface of the crystal until a point was found at which radio signals could be heard. The variable tunning capacitor allowed the operator to tune to various radio stations.
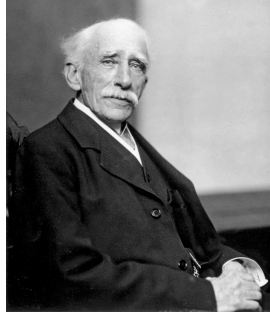
Figure 24 John Ambrose Fleming   (source: American Physical Society)

In 1904 John Fleming (Figure 24) invented the vacuum tube diode while working on various wireless detectors at the Marconi Company. The invention was based in part on work he had done on light bulbs in the 1880s at the Edison Company. The Fleming valve consisted of two elements enclosed in a glass bulb evacuated of air, a filament F and a plate P, as illustrated in Figure 25. An electrical current flowing from a battery through the filament heated the filament to a high temperature. The high temperature caused energetic electrons to escape from the filament and flow through the vacuum to the positively charged plate. The electric field within the tube prevented electrons from flowing in the opposite direction. Consequently, the Fleming valve performed the same diode function as the carborundum crystal. Conceptually the Fleming Valve receiver was the same as the crystal radio set shown in Figure 22 with the carborundum crystal replaced by the Fleming valve. The vacuum tube diode worked well as a detector, but the crystal detector worked as well and was much less expensive.
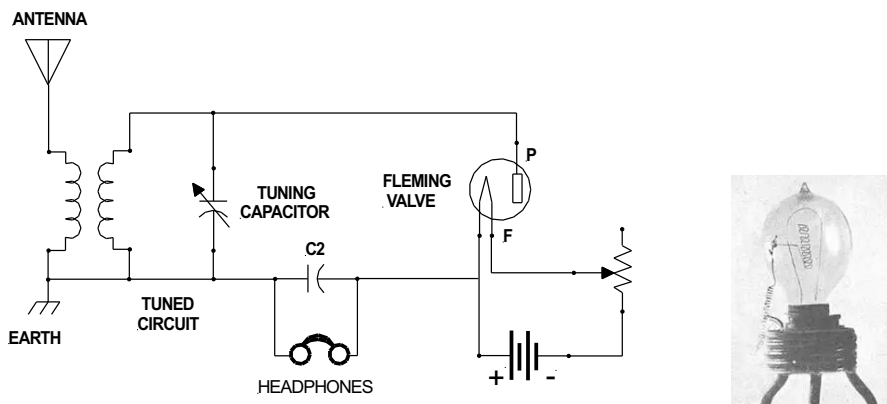

Figure 25   Fleming valve diode receiver. (source: EDN Magazine)

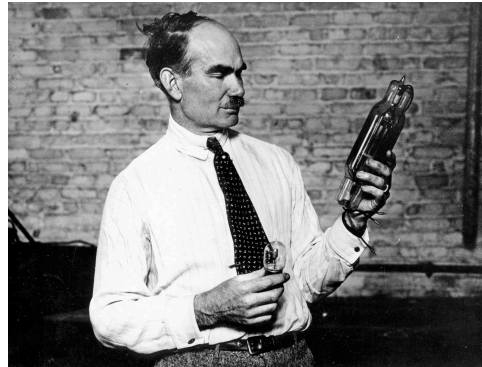**1.7   Beginning of the Electronic Age**

Figure 26 Lee DeForest   (source: gettyimages)

In 1906 Lee DeForest (Figure 26) invented the triode vacuum tube which he called the Audion. Many claimed that he had simply added a 3rd element to Fleming's diode. The courts finally concluded that DeForest's work was original. Fleming stated several years later that the idea of adding a third element to his diode had never occurred to him.

The three elements of an Audion vacuum tube were the filament F, plate P, and grid G enclosed in a glass bulb evacuated of air, as illustrated in Figure 27.
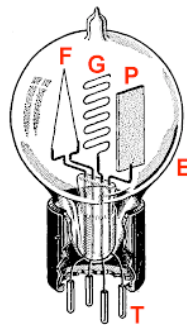

Figure 27 DeForest audion 3-element vacuum tube. (source: Wikipedia)

The audion worked the same as Fleming's valve with the exception that the grid controlled the flow of electrons from the filament to the plate. A weak input signal producing a small voltage change between the filament and grid resulted in a large voltage change between the filament and plate. The large plate voltage change was an amplified version of the input signal as illustrated in Figure 28.
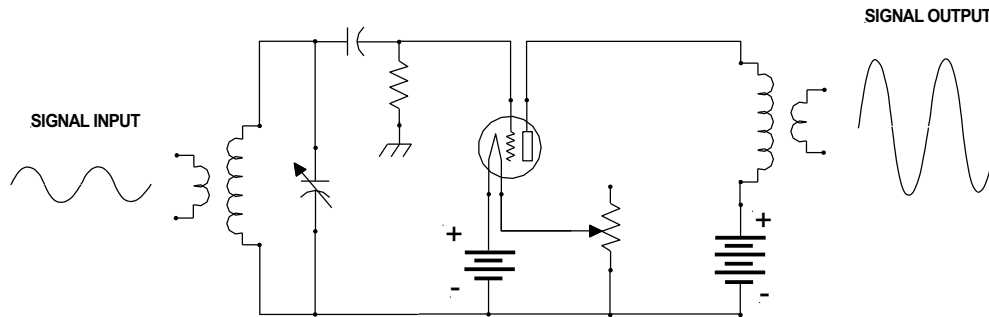
Figure 28   Audion amplifier circuit   (source: author)

The Electronic Age as we know it today began on October 26, 1906 when DeForest applied for a patent on "a device for amplifying feeble electrical currents". However, despite its enormous importance, the audion was basically ignored for five years or so.

Initially, no one, not evening DeForest, appreciated the importance of the audion as an amplifier. In 1914 AT&T (the telephone company) was committed to providing transcontinental telephone service and needed some type of audio amplifier. An audion modified for telephone operation was the answer. AT&T's success with the audion radically changed the wireless industry. Electronic circuits based on the audion quickly replaced the old spark-gap electro-mechanical devices producing new high performance electronic receivers and transmitters

Figure 29 Edwin Armstrong   (source: Radio Hall of Fame)

Edwin Armstrong (Figure 29) realized that considerable amplification could be achieved by feeding part of the amplified output signal back to the Audion's input causing the signal to be amplified over and over again. Armstrong used this concept to produce a simple highly affective regenerative receiver which he patented in 1914. An early regenerative receiver is shown in Figure 30. The regenerative receiver was extensively used by amateur radio operators for many years because of its simplicity and low cost. However, if not carefully adjusted, it would go into oscillation becoming a transmitter. At least some of the interfering noise on the amateur radio bands was caused by improperly tuned regenerative receivers.

Figure 30   Regenerative receiver   (source: Wikipedia)

A popular radio receiver in the 1920s was the Tuned Radio Frequency (TRF) receiver. This receiver, shown in Figure 31 and Figure 32, was simply a series of RF amplifiers tuned to the desired frequency followed by a detector and one or more audio amplifiers.

Figure 31 TRF Receiver   (source:Wikipedia)

Tuning a TRF receiver was a complicated process. To tune in a new station the three tuned circuit variable capacitors, on the left in Figure 32, had to be individually varied using the three large knobs on the front of the receiver. Tuning in a station was a process of successive approximation. Once a station was found, the numbers on the dials were written down, so the station could be found again.
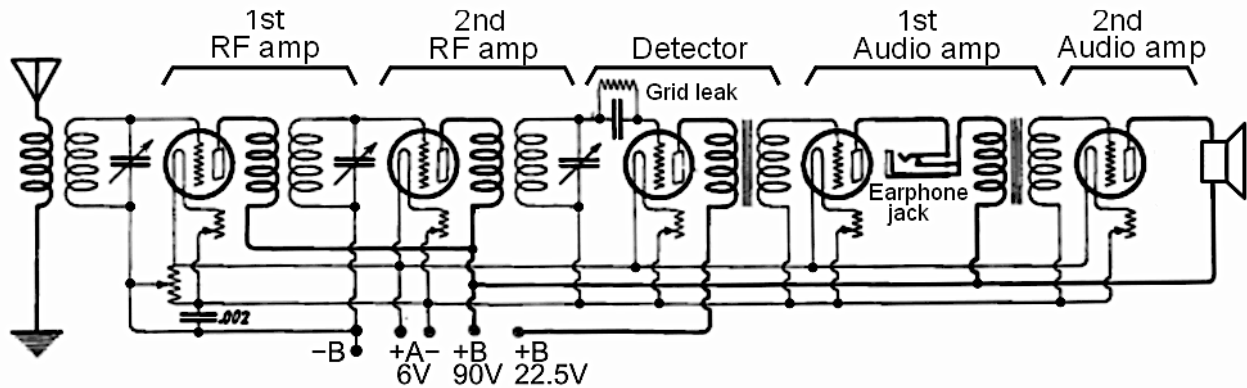
Figure 32 TRF Receiver Circuit Diagram   (source:Wikipedia)

In 1919 Edwin Armstrong developed a superheterodyne receiver that was considerably more sensitive and selective than both the regenerative and TRF receivers. The schematic of the receiver is shown in Figure 33.
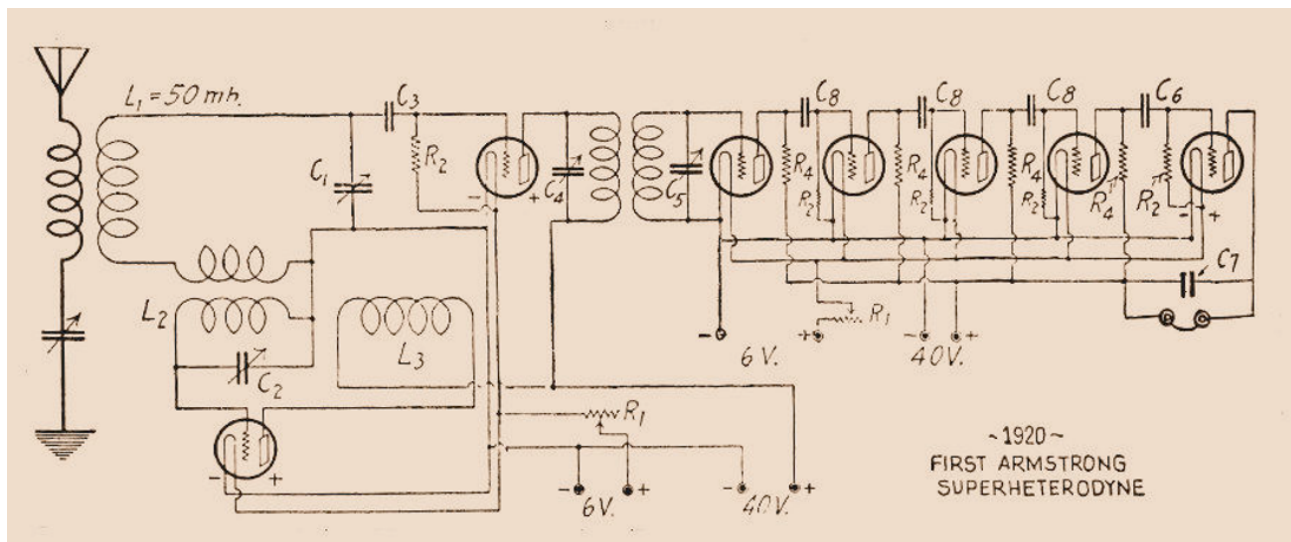

Figure 33 Armstrong's superheterodyne receivers   (source: Wikipedia)

The key feature of a superheterodyne receiver is mixing the incoming signal, entering the radio through inductor L_1, with a locally generated signal across inductor L_2. The locally generated signal is slightly lower frequency than the incoming signal. The mixing process produces a third resulting signal with a frequency equal to the difference between the incoming and locally generated frequencies. For example, if the incoming signal frequency is 1800 KHz and the local oscillator frequency is 1345 KHz, then the resulting frequency is 455 KHz. This resulting frequency is called the Intermediate Frequency or IF.

The capacitors C_4 and C_5 in parallel with the transformer primary and secondary windings in Figure 33 form a tuned circuit that allows only the 455 KHz IF frequency to pass through to the remaining receiver circuitry. The remaining circuitry is thus designed specifically for operation at 455 KHz. Because of its IF architecture the superhet design provides higher amplification and greatly improved selectivity compared to other receivers. Vacuum tubes 3, through 6 (counting from the left) are 455 KHz IF amplifiers. Tube 7 is the audio amplifier.

Tuning to a new station is simple. The frequency of the local oscillator is changed so that the difference between it and the frequency of the new station is 455 KHz. It is not necessary to know what the local oscillator frequency actually is. The operator simply tunes the local oscillator until the desired station is received.

The regenerative, TRF, and Armstrong's initial superheterodyne receiver all used a diode-capacitor peak detector, the same as used in the Figure 22 crystal radio. The peak detector was capable of receiving signals from spark gap transmitters and voice modulated radio waves. It was not capable of receiving CW signals produced by vacuum tube morse code transmitters discussed in the next section. This was a serious problem. The various waveforms of interest are shown in Figure 34. Operation of the peak detector is explained in Figure 35.

To solve the problem a Beat Frequency Oscillator (BFO) was added to receive CW morse code. A BFO is a local oscillator typically tuned 800 Hz below the IF frequency. The BFO and IF signals are mixed in a vacuum tube detector stage forming an 800 Hz audio signal that is passed on to the audio amplifier. The 800 Hz audio turns on and off in unison with the CW carrier forming an audible morse code signal which the operator can hear.
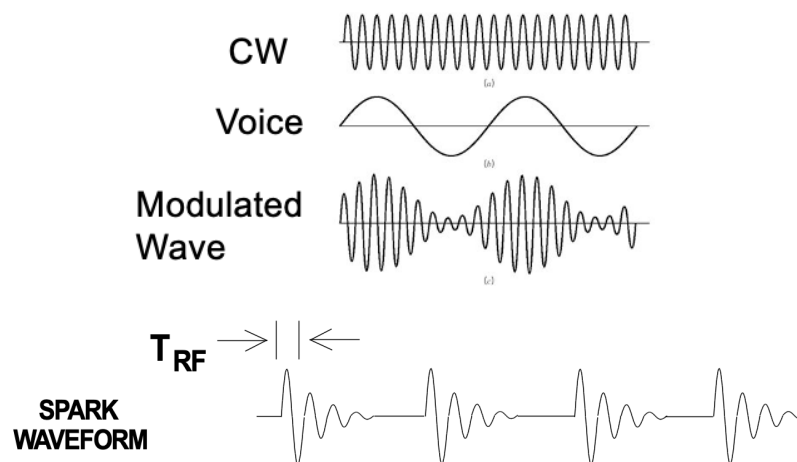

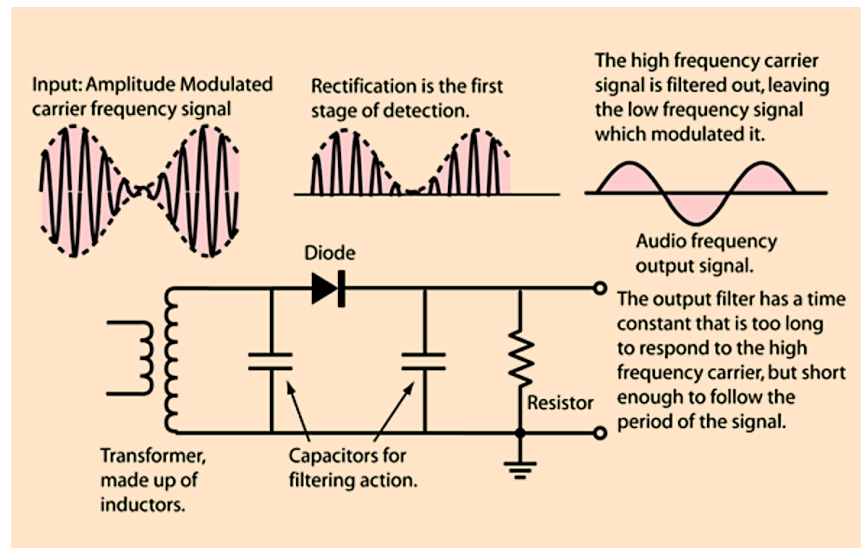Figure 34   Various Waveforms   (source: author)

Figure 35   Peak Detector operation   (source: HyperPhysics Concepts)

Adding a BFO to a high-end receiver, such as Armstrong's superhet, was not a problem. However, it was a problem for amateur radio operators trying to build low-cost equipment that they could afford. A vacuum tube CW morse code transmitter was far superior to the spark-gap transmitters that they had been using and well within their budgets. But a superhet receiver was costly. It was not an easy transition from spark-gap to CW for amateur radio operators. Various gimmicks were tried trying to receive CW morse code on their existing crystal radio and regenerative receivers. One attempt, for example, was to use a "chopper" consisting of a door bell buzzer to brake up their transmitter's continuous wave output into pulses that could be detected using existing low-cost receivers. Eventually, amateurs discovered how to modify single vacuum tube regenerative receivers to perform the BFO function in addition to receiving the incoming CW morse code signal. Simple two tube receivers were also designed in which one of the tubes was the BFO oscillator.

The receiver shown in Figure 33 is a single conversion receiver. Later more advanced receivers used dual and sometimes triple conversion eliminating some of the problems inherent in the single conversion radio while providing better sensitivity and selectivity. The newer receivers also provided better frequency stability, smoother more accurate tuning as well as the ability to "ride through" signal fades and noise crashes.

From 1925 on the superheterodyne design dominated the receiver market. The superheterodyne architecture has only recently been replaced by digital signal processing technology.

1.8   Transmitters

The problem of oscillation was not restricted to regenerative receivers. Audion amplifiers frequently broke into oscillation. However, it was quickly realized that with appropriate circuitry an Audion

could intentionally be caused to oscillate turning it into an RF oscillator capable of generating and transmitting continuous wave radio signals.

Figure 36 is the circuit for a Audion amplifier configured as an oscillator. The tuned circuit (inductor and variable capacitor) on the left determines the frequency of oscillator while the small inductor below it is the feedback circuit causing the amplifier to oscillate. The transformer on the right is the oscillator output.
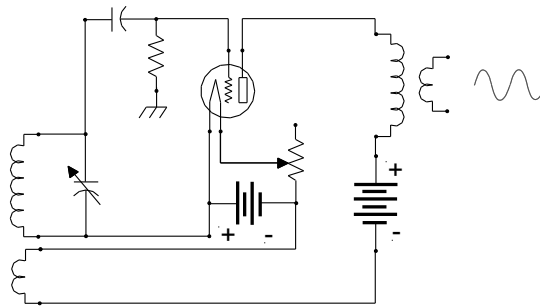
Figure 36 Audion oscillator   (source: author)

Figure 37 depicts a CW transmitter consisting of an oscillator stage followed by the output power amplifier. Notice that while operating the oscillator runs continuously while the morse code key forms dots and dashes by turning the output amplifier on and off.
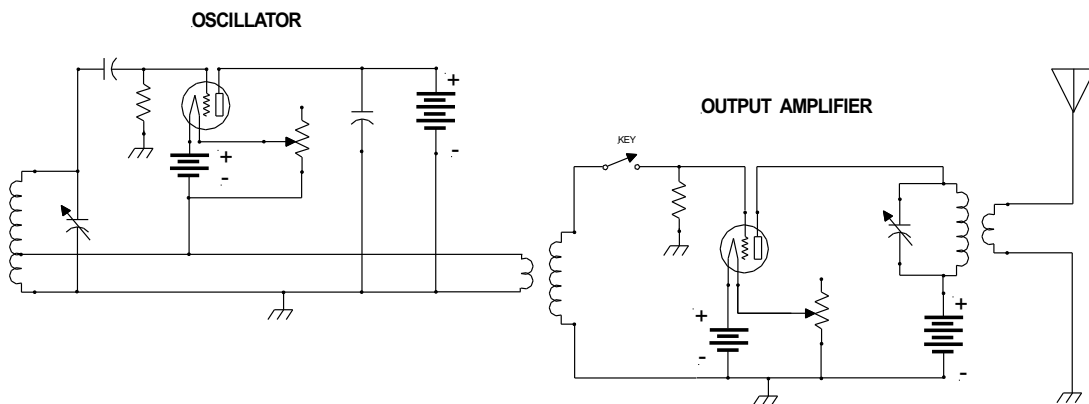
Figure 37 Early CW Transmitter   (source: author)

Figure 38 illustrates a CW transmitter typical of those used through the 1930s, 40s, and 50s. Adding a modulator to the final amplifier allowed audio to be transmitted permitting voice communications. The audio signal from the transmitter's microphone was amplified to a high-level and used to vary the plate voltage of the final amplifier. By changing the plate voltage, the amplifier's output power varied in unison with the microphone audio creating an Amplitude Modulated (AM) output signal

similar to that shown in Figure 34. This design was replaced in the early 1960s by CW – SSB (Single Side Band) transceivers utilizing transistor technology in place of vacuum tubes.
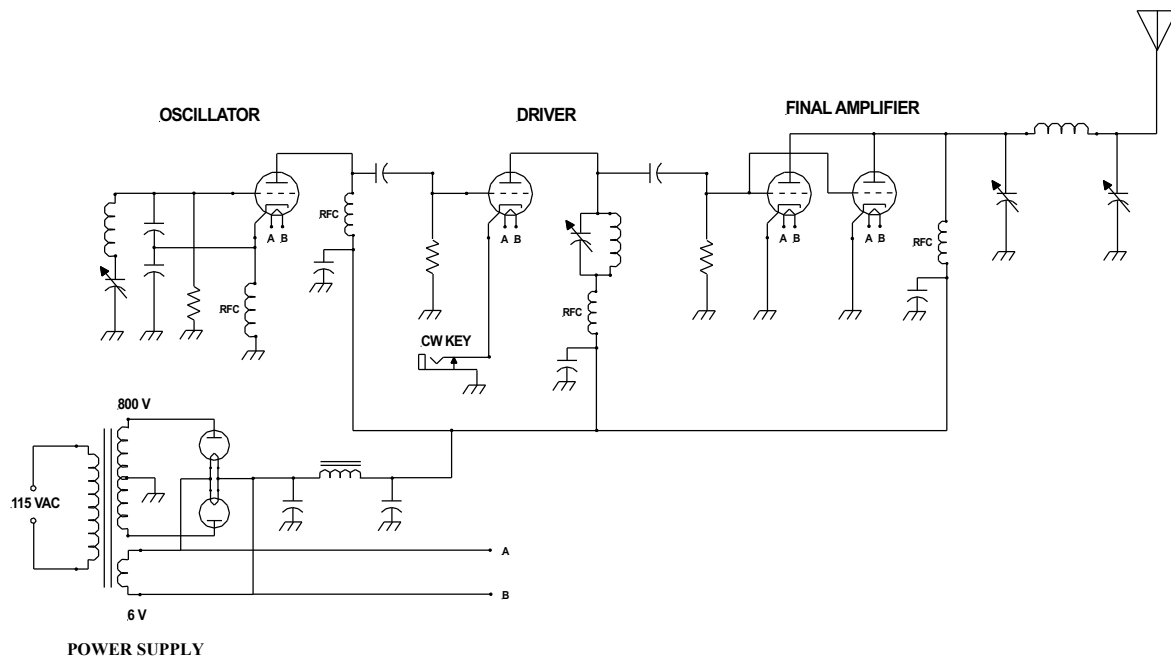

Figure 38 Typical CW Transmitter   (source: author)

1.9   Propagation Theory Did Not Match Results

In 1923, despite all that had been accomplished, how radio waves propagated was still unknown. From the mid 1990s through 1923 it was believed that radio waves traveled along the ground, or across the ocean surface. Complex theories were developed attempting to explain the long-distance propagation of radio waves. But the theories did not match the results of carefully conducted experiments.

In 1902 Oliver Heaviside (English) and Arthur Kennelly (United States) independently theorized that a conducting layer in the upper atmosphere could account for long range wireless transmissions. They suggested that wireless signals traveled back and forth between the conducting layer and earth's surface, first reflecting from the conducting layer, then reflecting from the earth as illustrated in Figure 39.
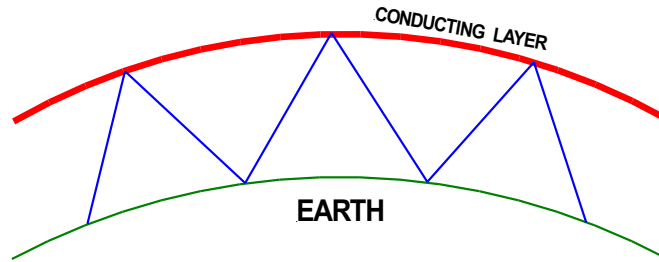
Figure 39   Presumed reflection of HF radio waves in Earth's ionosphere   (source: ResearchGate)

In 1910 English engineer and scientist William Eccles theorized that a body of electrons and positive ions existed in the upper atmosphere resulting from sunlight decomposing the air. He proposed that the level of ionization gradually increased as sunlight entered the upper atmosphere, reached a peak, and then slowly declined until it finally disappeared as illustrated in Figure 40. In this figure the black dots illustrate the level of ionization.
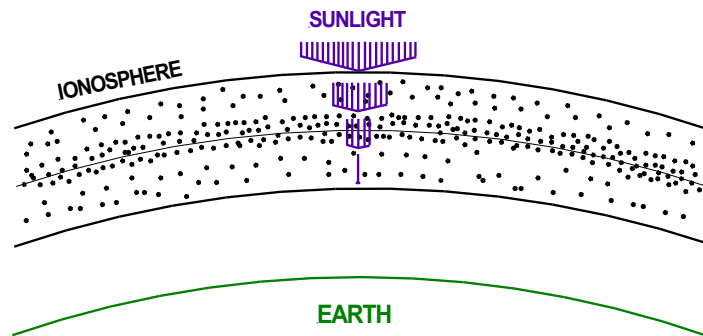

Figure 40 Ionization of Earth's upper atmosphere   (source: author)

Based on this concept, radio waves traveling upward into the atmosphere were refracted, bent back toward Earth, as they encountered increasing levels of ionization (Figure 41). The refraction of light at the boundary between two transparent medium was well understood. Eccles presumed that radio waves acted similarly. We are all familiar with refraction. Refraction creates rainbows and separates sunlight into its various red through purple colors as it passes through a prism (Figure 42).
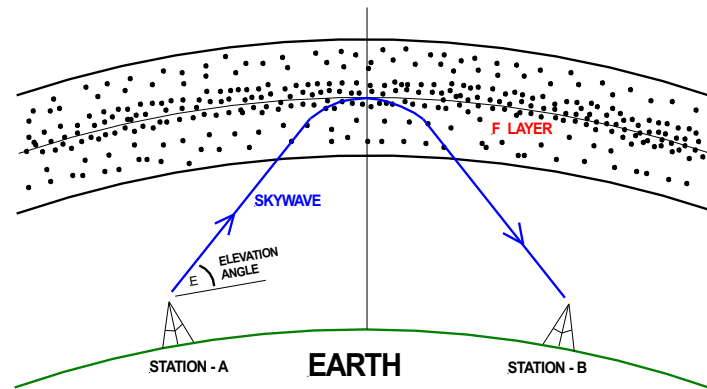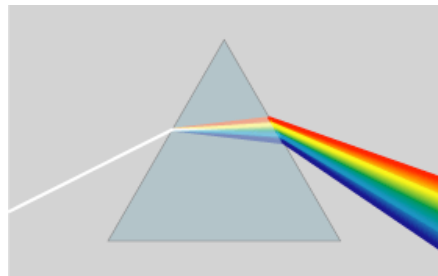
Figure 41   Refraction of radio waves in the upper atmosphere   (source: author)

Light refracting through a prism and through water droplets forming a rainbow of colors

Figure 42   Refraction of light   (source: Wikipedia)

But all of these ideas were simply theories. There was no proof that any of it was actually true.

In the early days of radio, amateur radio operators, with their home built wireless equipment, crowded the commercial wireless frequency bands. This was a serious problem from the perspective of commercial and military radio operators. To solve the problem legislation was enacted restricting the frequencies at which amateurs could operate. From 1912 on amateurs were exiled to the seemingly worthless short-wave frequencies of 200 meters and down, better known today as the

160, 80, 40, 20, and 10 meter frequency bands. While most amateurs operated at 200 meters (around 1.5 MHz), a few ventured down into the lower wavelength frequency bands. In the process they discovered the long-distance capabilities of "short wave" radio. Amateurs began communicating from the United States to Europe, New Zealand, and the Pacific rim on a regular basis using transmitters operating at relatively low power levels of 1,000 watts or less and small 20 meter antennas. In contrast, military and commercial companies operating in the 300 to 1,000 meter frequency bands required transmitters ranging in power from a hundred thousand to a million watts (Figure 37) and huge antennas (Figure 38) to cover the same distances.

Amateur radio operators contributed significantly to the understanding of long-distance radio communications. They discovered that short wave radio disobeyed the accepted theories of radio propagation. Short wave propagation, unlike long wave, seemed to involve only the atmosphere. Strange things happened at short wavelength frequencies. For example, it was possible to communicate with distant radio stations while closer radio stations could not be heard. Radio signals seemed to skip over the close by radio station creating what became known as the skip zone shown in Figure 43. We know today that a signal transmitted at a low elevation angle E travels further than a signal transmitted at high angles. It follows that relatively high elevation angles are required to communicate with close by stations. However, if the elevation angle is too high the signal will penetrate the ionosphere and be lost to outer space as is the case with the pink trace in Figure 43. The elevation angle at which penetration begins to occur is called the Maximum Usable Angle (MUA) illustrated by the red trace in Figure 43. Consequently, Station-B is the closest station that Station-A can communicate with. Stations between A and B can not be heard. They are skipped over. Why this occurred was not understood in the 1920s. At the time the skip zone was a complete mystery.
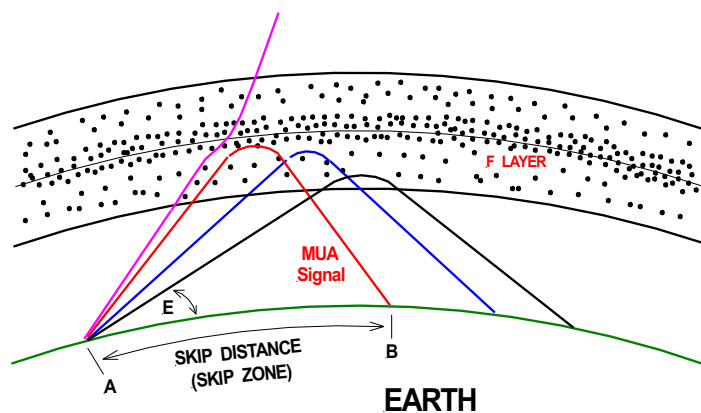

Figure 43   Skip distance   (source: author)

Figure 44 Edward Appleton (source: aip.org)

In April 1924 Edward Appleton studied the strength of radio signals received at Kings Collage Cambridge from the London BBC (British Broadcasting Company) radio station approximately 65 miles away. He discovered signal strength was constant during the day but varied at night. He theorized that at night he was receiving two signals, one traveled along the ground the other reflected from the upper atmosphere. Further investigations showed that radio waves were reflecting from atmospheric ionization about 60 miles above the Earth from what he called the electrified or E-layer. He later discovered that short wave radio signals passed through the E-layer and were reflected about 100 to 200 miles above the Earth. He named this electrified region the F-layer as illustrated in Figure 45. The signal absorbing D-layer was discovered later. Appleton's experiments proved that there were in fact ionized regions in the upper atmosphere that refract radio waves back to Earth.
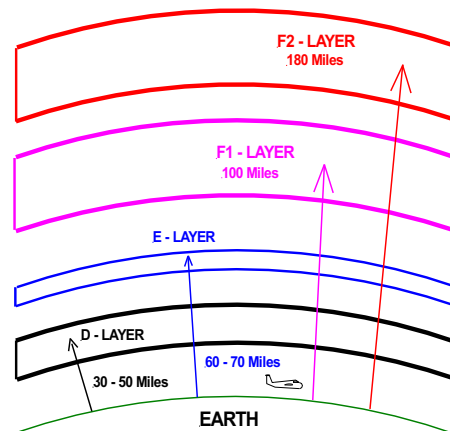

Figure 45 Layers of the ionosphere   (source: author)

Early on it was known that signal strength and quality varied throughout the day, and seasonally. Around the world short wave communications had been excellent and exciting since its discovery in

the late 1920s. Then early in the 1930s, for some unexplained reason, excellent long-distance communications disappeared. It was difficult to make any long-distance short-wave contacts. In the late 1930s long distance propagation conditions returned. It became apparent that HF signal strength and quality were varying with the solar cycle illustrated in Figure 46. HF radio propagation conditions were excellent in 1928 during solar maximum when large numbers of sunspots (Figure 47) were visible on the Sun. In 1934, during solar minimum, there were few if any sunspots resulting in poor HF communications. In 1938 solar maximum occurred again with large numbers of sunspots and excellent HF communications. By 1940 it was clear that skywave communications was directly affected by solar activity, particularly the 11-year solar cycle. Understanding the Sun and solar activity is thus critical to achieving reliable HF communications any where, any time, under any conditions.
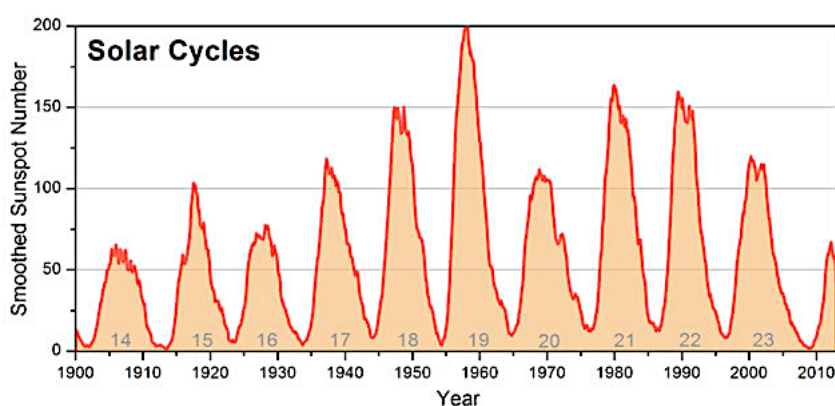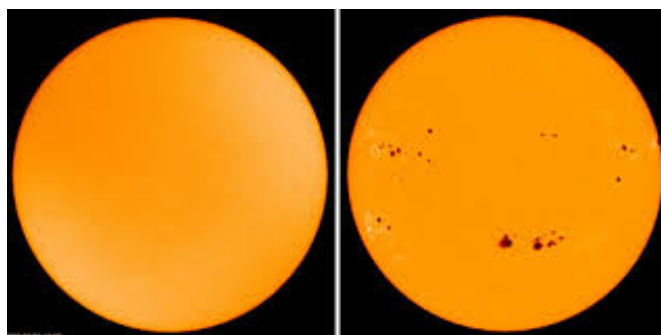

Figure 46   Solar sunspot cycle   (source: NOAA)

Solar Minimum   Solar Maximum

Figure 47   Sun at solar minimum and maximum (source: springer.com)

References

Yeang, Chen-Pang; "Probing The Sky With Radio Waves"; The University of Chicago Press, 2013

DeSoto, Clinton B.; "200 Meters & Down" The American Radio Relay League, 1936

McNamara, Leo F.; "The Ionosphere: Communications, Surveillance, and Direction Finding"; Krieger Publishing Company, 1991

Davies, Kenneth; "Ionospheric Radio"; Peter Peregrinus Ltd.,1990

Hunsucker R. D.; Hargreaves, J. K.; "The High-Latitude Ionosphere and its Effects on Radio Propagation"; Cambridge University Press 2003

Nichols, Eric P.; "Propagation and Radio Science"; The American Radio Relay League, Inc. 2015

Aitken, Hugh G. J.; "Syntony and Spark, The Origins of Radio" Princeton University Press, 1985

Aitken, Hugh G. J.; "The Continuous Wave Technology and American Radio, 1900 - 1932" Princeton University Press, 1985

Bucher, Elmer Eustice; "Practical Wireless Telegraphy" Wireless Press, Inc, 1917

Collins, Archie Frederick; "The Radio Amateur's Hand Book" 1922

Collins, Archie Frederick; "Wireless Telegraphy" McGraw Publishing Company 1905

Coe, Lewis; "Wireless Radio A History" McFarland & Company Inc, 1996

Trowbridge Jr., Calvin D.; "Marconi" Calvin D Trowbridge Jr., 2009

Devoldere, John; "Low-Band DXing" fourth edition; ARRL, 2005

Levis, Curt A. ; Johnson, Joel T.; and Teixeira, Fernando L.; "Radiowave Propagation Physics and Applications"; John Wiley & Sons, Inc., 2010

Goodman, John M.; "Space Weather & Telecommunications"; Springer Science+Business Media Inc. 2005

Forbes, Nancy and Mahon, Basil; "Faraday, Maxwell and the Electromagnetic Field" Prometheus Books 2014

Nahin, Paul J.; "Oliver Heaviside" The John Hopkins University Press 2002

Stross, Randall; "The Wizard of Menlo Park" Three Rivers Press 2007

Moyer, Albert E.; "Joseph Henry" Smithsonian Press 1997