(2)

NAVAL POSTGRADUATE SCHOOL
Monterey, California

AD-A257 560

THESIS

92-30445

A Technical Review of Cellular Radio and Analysis of a Possible Protocol

by

William David Reese

September 1992

Thesis Advisors: Dr. G.M. Lundy
Lou Stevens

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Computer Science Dept. Naval Postgraduate School; 6b. OFFICE SYMBOL (if applicable) CS/Ln	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000	7b. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION; 8b. OFFICE SYMBOL (if applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
A TECHNICAL REVIEW OF CELLULAR RADIO AND ANALYSIS OF A POSSIBLE PROTOCOL

12. PERSONAL AUTHOR(S)
Reese, William D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Master's Thesis	FROM 04/90 TO: 09/92	September 1992	86

16. SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the United States Government.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Radio, Television, AM, FM, Cellular Radio, Cellular Radio Protocol

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Radio and Television technology made possible the field of computer science. This thesis shows the development of radio and television technology from both an historical and technical aspect. A review of the important researchers and their contributions is followed by a technical explanation of the theories behind electromagnetic radiation of radio and television signals and the technology which was developed to implement such transmissions. The evolution of development which the paper outlines begins with the some of the first theories about electricity and magnetism and the subsequent mathematical foundation developed to explain them. This is followed by a number of experimental and developmental researchers and their contributions.

The bulk of the paper is concentrated on explaining the earliest generations of radio and all generations of television. The major components of both radio and television are described in detail along with an explanation of what they do and how they work. Such components, in many cases, found important uses in fields outside those for which they were developed. A brief overview of the regulatory environment of each technology and the U.S. and international standardization efforts is also included.

Finally, the paper illustrates a modern-day application of radio technology--the cellular radio industry. A description of the components and their functions is followed by a possible cellular radio protocol and analysis.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT [X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT. [ ] DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL G. M. Lundy	22b. TELEPHONE (Include Area Code) (408) 646-2094; 22c. OFFICE SYMBOL CS/Ln

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

Approved for public release; distribution is unlimited

A Technical Review of Cellular Radio and Analysis of a Possible Protocol

by
William David Reese
Captain, United States Army
B.S., Colorado State University, 1980
M.S., Colorado State University, 1989

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 1992

Author: William David Reese

Approved By: Professor G.M. Lundy, Thesis Advisor

Professor Lou Stevens, Co-Advisor

Professor Robert B. McGhee, Chairman,
Department of Computer Science

ABSTRACT

Radio and television technology made the field of cellular radio possible. This thesis shows the development of radio and television technology from both an historical and technical aspect. A review of the important researchers and their contributions is followed by a technical explanation of the theories behind electromagnetic radiation of radio and television signals and the technology which was developed to implement such transmissions. The evolution of development which the paper outlines begins with some of the first theories about electricity and magnetism and the subsequent mathematical foundation developed to explain them. This is followed by a number of experimental and developmental researchers and their contributions.

The bulk of the paper is concentrated on explaining the earliest generations of radio and all generations of television. The major components of both radio and television are described in detail along with an explanation of what they do and how they work. Such components, in many cases, found important uses in fields outside those for which they were developed. A brief overview of the regulatory environment of each technology and the U.S. and international standardization efforts is also included.

Finally, the paper illustrates a modern-day application of radio technology--the cellular radio industry. A description of the components and their functions is followed by a possible cellular radio protocol and analysis.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

This paper examines the development of radio and television from both a historical and technical perspective. A discussion of the significant individuals and the technology they developed is included. The questions of how and why these technologies worked should be answered in the pages that follow. After a technical and historical discussion of the technology involved, a modern day application, cellular radio, is illustrated along with a sample protocol.

Wireless communication has a history that predates the development of the radio considerably. The ancient Greeks are known to have used a series of beacon fires to relay messages over great distances. Natives of the African Congo used drums to send short messages related to war, hunting, sickness, and death over much longer distances than possible with the voice. On this continent, the American Indian used smoke signals to relay brief messages across vast stretches of open prairie and dessert.

More complex messages were also sent quickly across vast distances hundreds of years before the development of radio. As early as the 17th century, the British had developed a system of communication by semaphore. It involved hoisting a series of shapes, each with a particular meaning, to the top of a tower in a particular sequence. The towers allowed the shapes to be recognized from great distances where, in turn, they could again be hoisted atop the distant towers to further extend the range of the message. Indeed Napoleon set up a series of 1200 semaphore towers stretching from Paris to the Russian front!

These methods, while effective, were cumbersome and slow--especially when great distances were involved. The telegraph solved the problem of speed when it came on the scene in the 1830's. However, the telegraph had its limitations--primarily the need for expensive wire connections. The telephone had a similar limitation. Fortunately for those who have neither the resources to arrange a wire connection or perhaps the possibility (ships, planes, and vehicles), the radio was developed. Radio broke the bonds of wire while retaining the speed and range of other systems.

The development of radio built upon the technology of the telegraph and telephone. Radio research and development, in turn, laid the foundation for advances in many other areas of communication--most specifically, the television. Perhaps of most significance is the array of technical advances in electronics that radio-related research brought about. It was the desire to improve upon radio reception that brought about the triode (a vacuum tube used in early amplification). This, in turn, was improved upon in the 1940's when the transistor was invented. The integrated circuit, of course, came soon thereafter. All of these technologies assisted in the creation and advancement of computer hardware.

One of the fastest growing applications of radio technology is that of the cellular telephone. Advances in low-power radio transmission and the speed with which modern computers can aid in frequency management and signal processing have made cellular radio a commercial reality today. This paper will briefly illustrate a how cellular radio would operate and analyze a possible protocol which could be used.

II. HISTORICAL DEVELOPMENTS

The development of radio and wireless communication began over 100 years ago in Europe. It has become an ever-expanding field of study, research, and commercialization. From its inception, wireless communication has grown to include the transmission of voices, pictures, and vast amounts of data over ever increasing distances.

A. EARLY RADIO

The earliest forms of radio were quite unlike their modern day counterparts. Indeed, the developers of the first radios had no preconceived notions about using such devices to communicate. Once the idea of using radio as a means of signalling came about, the radios of the day were made to become "wireless *telegraphs*," transmitting "dots and dashes" through the air. Several years passed before "wireless telephony" was developed and voices, music, and other types of sound could be transmitted via radio.

1. The Pioneers

The first developers of radio built upon many theories developed in the 19th century.

a. Michael Faraday

Wireless communication can trace its modern-day roots to Michael Faraday (1791-1867), the famous English chemist and physicist. The child of a blacksmith, Faraday lacked a rigorous formal education. He went to work for a book binder at the age of 13 and, at 19, was able to accompany his employer to a series of lectures on natural philosophy. Faraday was charged with making notes of the lectures that would subsequently be bound into books that described the experiments of the various lecturers. It was during this period (age 21) that he became fascinated with science and he sought employment as an assistant to a lecturer at the Royal Institution in London. He subsequently worked his way up to lecturer and devoted ever increasing amounts of time to experiments. Although he had become an accomplished chemist, his research in the 1830's was devoted almost entirely to the study of electricity. It was in 1831 that Faraday discovered how to produce electricity from magnetism. Lacking a formal mathematical background, Faraday was unable to explain many of the electromagnetic phenomena that his experiments produced. Nevertheless, his work in the field is considered to be classic and his theories concerning the existence of electromagnetic fields were quite controversial and rejected by most scientists of his day.

b. James Clerk Maxwell

The electromagnetic theories of Faraday remained controversial until the Scottish physicist and mathematician James Clerk Maxwell (1831-1879) was able to reinterpret Faraday's ideas within a solid mathematical framework. In 1865 Maxwell published "A Dynamical Theory of the Electromagnetic Field" which outlined the basis for his subsequent work in the area. It was in Maxwell's famous treatise Electricity and Magnetism where a mathematical explanation for Faraday's electromagnetic wave theories was presented. Known as the famous Maxwell equations, these mathematically derived formulae were able to support Faraday's ideas. Among other things, Maxwell's equations indicated that the flow of an electric current would generate electromagnetic radiation. (Chapter III includes a more detailed discussion of Maxwell's equations.) Since Maxwell's work was almost exclusively theoretical in nature, the actual production and detection of these electromagnetic waves was never incorporated into his work.

As early as 1837, electromagnetism had made possible the early versions of the telegraph. Unlike wireless communication, the telegraph relied on a physical wire to comprise most of the circuit of an electromagnet. The operator of the telegraph key manually closed or completed the circuit each time the telegrapher's key was depressed.

This closing of the circuit caused the electromagnet at the receiving end to develop its magnetic field. By noting in what order and for how long the field was active, the operator at the receiving end of the telegraph circuit was able to determine the message being transmitted using Morse code or some other protocol. These electromagnetic signals were often made to move magnetic needles, produce an audible signal, or create a permanent record at the receiving end [FREE58]. Radio (and television) would eventually borrow from and build upon the technology developed by the telegraph industry by applying many aspects of telegraphy to a new wireless mode of communication. But before wireless communication could evolve, experimental confirmation of Maxwell's electromagnetic theories was necessary.

c. Heinrich Hertz

It was not until 1888 that Maxwell's theory of electromagnetic radiation was confirmed experimentally. During that year the German physics professor Heinrich Hertz (1854-1894) successfully measured the velocity of radio waves in his laboratory. Professor Hertz generated the radio waves by generating a string of sparks across the secondary winding of a transformer (a device used to change voltages or, in this case, to induce voltage into a circuit without being directly connected to it [SCRO71]). As the sparks jumped the gap between the metal spheres (which, in turn, were connected to terminals of the transformer's induction coil), the electromagnetic impulses were created. Hertz suspected that radio waves would behave in many ways like light waves. By setting up a sheet of metal a few meters away, he was able to reflect the electromagnetic waves back in the direction from which they came. By adjusting the distance between the transmitter and the reflecting sheet of metal, Hertz was eventually able to create a "standing wave" (a wave reflected back along the same path it came from) that he subsequently used for wavelength measurements. These electromagnetic waves could be detected using a circular loop of wire with a small gap (~ 1 micrometer). In the darkened laboratory, these waves formed sparks in the loop of wire's small gap as they passed through. Additionally, Hertz demonstrated that the waves could be reflected with a sheet of flat or curved metal in the same way light could (the wavelengths originally used by Hertz eventually became widely used for modern radar applications). Further experimentation and measurement led Hertz to conclude that the speed of these waves was approximately 186,000 miles per second--the speed of light. Detecting these waves at greater distances proved to be a greater challenge that Hertz was unable to pursue (he died of blood poisoning in 1894 at the age of 36). Today the term Hertz (or Hz) is used extensively in electronics as a synonym for cycles per second. It was the description of Hertz' experiment in an electrical journal that caught the eye of Gugliemo Marconi, one of the greatest scientist-entrepreneurs of the era, that subsequently brought the world into the age of wireless communication.

d. Edouard Branly

The first long-distance detector of wireless waves was the coherer, invented by Edouard Branly, a French physicist. Branly borrowed the basis for the coherer from medical science[DUNL44]. After learning that nerves, which carry signals from the skin to the brain, are not continuous fibers but instead closely massed neurons, he applied this principle to the coherer. In 1891 Branly filled a glass tube with loose iron filings, plugged the tube with metal plugs, and then placed the tube in a closed circuit with a battery and a galvanometer. When electromagnetic or "Hertzian" waves were produced 25 yards away, the iron filings cohered together (hence the name "coherer") and completed the closed circuit. To be useful in transmission of data, the coherer had to be "tapped" after each detection of a signal in order to disperse the cohered particles. Although it was incapable of demodulating signals (it could only detect the presence or absence of a signal), the coherer worked well for the on-and-off (digital) type of signals involved in Morse code transmissions and remained the centerpiece to radio reception for many years.

e. Gugliemo Marconi

With all the ingredients in place for the development of wireless communication, Gugliemo Marconi(1874-1937), the son of a wealthy Italian-Irish couple, set his mind to using the Hertzian waves to transmit messages. While on vacation in 1894, Marconi read an article in an electrical journal describing Hertz's work and its implications for other possible uses. He immediately cut short his vacation to return to Italy and develop his first radio. Building upon the design of the Hertz transmitter, he made several modifications to adapt its use to broadcasting messages. Perhaps his most clever initial modification was to incorporate a telegraph key into the circuit. This allowed the user to control the duration of each broadcast of electromagnetic waves--dots and dashes via wireless! He also greatly increased the range of the transmissions by grounding (connecting) half of the Hertz oscillator to the earth and elevating the other half (extending via an antenna) into the air. By June of 1896, he had been granted a patent in Britain for a single-circuit transmitter & single-circuit receiver.

By 1897, Marconi's success had so impressed British capitalists that major financial backing was offered for the inception of what was to become the dominant wireless communications company[ARCH38]. Marconi's Wireless Telegraph Company, Ltd. (also known as the British Marconi Company) became a massive industrial empire over the next two decades.

Marconi spent the rest of his life improving upon this new means of communication. By March of 1899 his signals had been sent across the English Channel. He successfully transmitted *signals* across the Atlantic in December 1899 (it was not until 3 years later that *messages* were sent across the Atlantic). Commercial transmission services across the Atlantic began in 1907. The adaptation of radio as a modern means of communication is due more to Marconi than anyone else. Other, seemingly obvious, candidates to commercialize radio (such as the telegraph & cable industry, the telephone industry, or the electrical manufacturing industry) all seemed oblivious to the potential of this discovery.

Marconi did not consider himself a scientist. In fact his educational background was comprised of private tutoring at home as a child. He did, nevertheless, possess an extreme interest in science along with a vivid imagination. These traits, combined with the financial resources of his family and his own patience and diligent work, eventually won him the Nobel prize in physics as well as the great fortune and fame that accompanied his commercialization of radio.

f. Alexander Popoff

The Russian pioneer in wireless communication was Alexander Popoff (1859-1906). In an attempt to detect thunderstorms in advance, he conducted numerous experiments using the coherer developed by Branley. He wanted to use the Branley coherer as a detector of the static electricity in the air to detect the approach of thunderstorms. Eventually moving on to wireless signal applications, Popoff concentrated on increasing the distances over which Hertzian transmissions could be detected. By adding, among other things, a piece of wire that stretched from the air above the coherer, through the coherer, into the ground, Popoff created an early antenna and was able to demonstrate reception of signals over 600 yards in 1895. By 1899, he was able to increase the range to 45 miles. Much of Popoff's work paralleled Marconi's in time and substance. He arguably invented or improved upon several aspects of early radio before Marconi did. However, because of his reluctance to commercialize his ideas or seek patents for them, his impact on early radio and antenna development is often overlooked.[DUNL44]

g. Oliver Lodge

In the late 1890's the English physicist Oliver Lodge found ways to "tune" the Hertz transmitter or receiver through the use of an adjustable induction coil. Like Hertz, Oliver Lodge had been conducting experiments on "standing" electromagnetic waves at about the same time as Hertz. Unlike Hertz, Lodge had created his stationary waves within the confines of wires and not in open space. Perhaps because of the untimely death of Hertz in 1894, it was Oliver Lodge who became known for his work in the tuning of radio transmitters and receivers.

Lodge's early work included an investigation into lightning conductors in 1885. Since the 1870's, lightning rods had been used extensively in Britain in an attempt to prevent damage to the vast network of telegraph lines during thunder storms. Lodge was able to differentiate between what was believed to be the flow of direct electrical current from the clouds to the ground and the *alternating* current that actually comprised lightning strikes. As predicted by Maxwell's equations, alternating currents behaved quite differently than direct currents which flowed at a constant speed and direction. Lodge's work on lightning rods led him to a thorough understanding of the differences and relationships between inductive and capacitive reactance in a circuit[AIT76--pp.87-88]. (In both cases, the reactance is a function of the frequency at which a current is changing direction. A segment of a circuit such as a coil of wire, with high inductance, would offer little resistance (impedance) to a low frequency alternating current and high resistance to a high frequency current. The opposite holds true for the portion of a circuit high in capacitance.)

By properly arranging the inductance and capacitance in a circuit such as a radio transmitter or receiver, Lodge realized that a circuit could be tuned (or resonant) at certain frequencies. The early versions of transmitters (those constructed by Hertz, Marconi, and others) were not tuned at all. They simply transmitted their signal in a very broad band of frequencies simultaneously. While this made the tuning of receivers less necessary, it did have the distinct disadvantage of wasting bandwidth. Two or more transmitters operating in the same effective range would interfere with each other's signals causing both to turn into garbage at the receiving end. Oliver Lodge quickly grasped the importance of tuned radio circuits.

Lodge had discovered a fundamental trade-off between "open" circuits (such as the spark gap generator of Hertz design) and "closed" circuits (those whose inductance and capacitance had been carefully designed to be resonant or tuned). The open circuits were very efficient at radiating electromagnetic waves into space. However, since they lost energy (dampened) rapidly, the oscillations quickly died down--generating waves of differing frequencies in the process. Closed circuits, on the other hand, were poor radiators of electromagnetic signals but their extremely slow damping produced waves of a much more consistent frequency. Lodge devised several techniques for overcoming this conflict. His first idea was to abandon spark transmissions and generate continuous waves using alternators to generate the alternating current required. Unfortunately the generators of that day were designed for relatively low frequencies (50 or 60 Hz)--much lower than radio frequencies. Lodge was forced to compromise by combining the open and closed circuit designs to achieve sub-optimal tuning and sub-optimal radiating capability. This was done by focusing his attention on the antenna systems of both the transmitter and receiver[AITK76]. In 1897 he received a patent for his tuning mechanism which became tremendously important as the airwaves became more and more crowded. This key patent was subsequently acquired by the Marconi Company.

h. Reginald Fessenden

A key area of research in radio transmission at the turn of the century involved voice communication. Radio telegraphy had proven itself very valuable but, perhaps because the telephone had become commonplace, the pursuit of voice radio seemed logical. In 1901 Professor Reginald Fessenden of the University of Pittsburgh was able to transmit and receive faint voice tones. Fessenden's use of the spark gap type transmitter was really not satisfactory for voice transmission. The dots and dashes of Morse code could easily be sent out via discrete electromagnetic wave bursts of various duration. However, for continuous voice waves, a continuous transmission of electromagnetic waves seemed necessary. While Fessenden was able to introduce voice signals to the antenna, the continuous din of the spark transmitter's transmission almost completely drowned them out. The problem was basically that the frequency of the spark transmitter was relatively low--within the range of audible frequencies (20-20,000 cycles per second). By increasing the frequency of the spark transmitter as much as (then) technically possible (to around 10,000 cycles per second) he had limited success with the transmission of voice signals. Unfortunately the tones were very faint due to the alternators (devices used to convert a continuous power supply into one that "alternates" its positive and negative current at high frequency) available at the time. As ever higher frequency alternators were developed, the strength of the signal increased as the frequencies used moved above the audible range. Finally, on Christmas Eve, 1906, after the development of the first 80,000 cycle alternator, a successful broadcast of both voice and music was made to ships & stations within a several hundred mile radius of the broadcast site.[ARC26] Fessenden's work in continuous wave transmission also led him to improve upon the coherer-based reception techniques of the day by inventing the heterodyne receiver in 1905. The heterodyne used the difference in frequency between transmitter and receiver to produce an audible tone at the receiving end. The technique required the receiving circuit to generate its own frequency in order to combine it with the incoming frequency. If the resulting difference in frequencies was within the audible range (20 to 20,000 cycles per second), the heterodyne transformed the signal received into one that the listener could hear. This technique proved very cumbersome with the available technology and delayed the acceptance of the heterodyne receiver for several years.[MACL49].

i. General Henry Dunwoody

Radio continued to improve through the years as various improvements and innovations were made. Among them were the invention of the crystal receiver in 1906 by the U.S. Army General Dunwoody. He discovered that carborundum, a product of electric furnaces, made an excellent detector of electric currents. (Detecting such electric currents coming from a receiving antenna was key to determining if radio signals were present.) Other substances such as lead ore, iron pyrites, and silicon were also found to have this property at about the same time.[ARCH38]. The efficiency and low cost of the crystal receiver soon made the older coherer obsolete while bringing the cost of radio receivers within range of thousands of amateur radio enthusiasts.

j. Lee DeForest

The 1906 development of the triode (also known as the audion) by Lee DeForest made voice amplification possible. In 1884, Thomas Edison had experimented with a similar device while trying to develop electric light. Known as a two-electrode light, Edison noticed that a current would flow only one way through it--like a one way valve. Twenty years later Dr. J. Ambrose Fleming, a technical advisor to Marconi, proposed its use to step down radio frequencies received into the audible range. Fleming realized that this one-way valve could be used to convert an alternating current into a direct current (since the valve only permits the flow of current through it in one direction). This 'Fleming valve' became one of the earliest vacuum tube technologies to be used in radio.

In 1906, Lee DeForest modified the two-electrode tube further by placing a tiny grid iron, comprised of very fine wires, between the base plate and filament. This modification had the effect of amplifying or increasing the flow of current between the filament and plate by three fold. (The addition of this third element caused the device to be known as the "triode" or DeForest "audion tube.") Further, several of these devices could be used in series, creating even greater amplification of the current.

The significance of this invention convinced DeForest to establish the DeForest Radio Telephone Company in order to directly exploit his invention. A major procurement of DeForest equipment in 1907 by the U.S. Navy proved a tremendous help in getting the company off the ground. Unfortunately for DeForest, his company suffered in the aftermath of the Marconi stock scandal of 1912. (After a sharp run up in the price of Marconi stock, the price dropped dramatically when it was discovered that British government officials, who had contracting authority for the purchase of Marconi equipment and services, had profited by their inside information.) Disappointed investors successfully called for the prosecution of DeForest for allegedly defrauding them with respect to their investments in the DeForest Radio Telephone Company. The bad publicity created dire financial circumstances for DeForest after his release on bail. It was during this period that AT&T Co. arranged to purchase the wire-line rights to the audion patent--for use in long distance telephone service, at the bargain price of $50,000 in 1913. [ARCH38]

k. Edwin Armstrong

One of the most stubborn problems to overcome with radio involved natural and man-made static. Professor Edwin Armstrong of Columbia University did a great deal of research in this area. By 1933 he had perfected a system of frequency modulation (FM) to largely eliminate the problem of static. As a major shareholder in the RCA Corporation, he was able to convince RCA to support his research into this area. With the resources of RCA's laboratories and technicians, Armstrong became convinced that FM was a viable form of radio worthy of further development and superior in quality to AM. Armstrong was devastated when RCA decided not to pursue further development of FM in 1934. Television technology had been developing in laboratories since the 1920's. RCA was nearing a series of major breakthroughs in the area of television and was concerned that the commercialization of FM would complicate the expansion of TV by using up valuable bandwidth in the electromagnetic spectrum. Eventually, Armstrong was successful in building his own experimental FM station which became operational in 1938. The superiority of FM broadcasts became instantly obvious and, by 1944, there were 52 FM stations and over half a million receivers capable of receiving FM broadcasts.

2. Time line of Radio Evolution

To put the development of the modern radio into perspective, a few of the milestones in its development follow.

a. First Radio Transmissions

While it could be argued that Hertz had actually transmitted radio *signals* in his laboratory in 1888, the first to utilize such radio transmissions to convey messages was Oliver Lodge at Oxford in September of 1894.[AITK76] (There appeared for many years to be a controversy over whether Lodge or Marconi was actually first.) Unfortunately for Oliver Lodge, he was more of a scientist than an entrepreneur and his early radio work was quickly overshadowed by the activities of Marconi. It was Marconi who rapidly commercialized the idea and received the first patent for wireless telegraphy in England.[ARCH38]

b. Crossing of the Atlantic

In December of 1899, Marconi successfully transmitted and received signals sent via radio across the Atlantic using the protocol developed by Samuel Morse for the telegraph. This first demonstration involved only the transmission of the letter 'S' repeatedly. Commercial trans-Atlantic service began seven years later.

c. Voice Transmissions

Professor Fessenden's first voice transmissions took place in 1901. It took another five years for the appropriate equipment to be developed to make long-range transmission possible.[ARCH38]

d. Television and FM

FM came to be largely as a result of the efforts of Professor Armstrong in the 1920's and 1930's. The first experimental FM broadcasting station was operational in 1938 and was quickly followed by numerous commercial broadcasting stations. Ideas and proposals for "electric pictures" became widespread during the 1880's.[SHIE77] Television's development began in earnest during the 1920's. Although public demonstrations on early television had occurred as early as 1926, television was not commercially developed until the late 1930's. The 1950's brought color television and the resolution and quality of these "electric pictures" continues to improve to this day. Technology contributed by the development of the television also proved quite valuable in the development of the videophone or "picture telephone" of the 1990's.

TABLE 1: TIMELINE OF RADIO MILESTONES

Year	Event	Person
1865	Formal Theories of Electricity and Magnetism Published	Maxwell

TABLE 1: TIMELINE OF RADIO MILESTONES

Year	Event	Person
1888	First Radio Transmissions	Hertz
1894	First Radio Messages Transmitted	Lodge
1899	First Trans-Atlantic Radio Transmission	Marconi
1901	First Voice Radio Transmissions	Fessenden
1906	Invention of the Audion--making amplification possible	DeForest
1906	Discovery of low cost crystals for radio reception	Dunwoody
1909	First 100,000 Hz Alternator developed for radio transmission	Alexanderson
1920	First (licensed) AM Broadcasting Stations Operational	Various
1938	First FM Broadcasting Station Operational	Armstrong

B. MODERN RADIO

The evolution of radio continues today. As outlined in the following section, the technological advances in the areas of radio components have enabled the radio to become cheaper, smaller, more powerful & reliable, and, as a result, more common. Radio has evolved into a world-wide commercial industry that impacts the majority of the world's population as a primary media source.

III. TECHNICAL DEVELOPMENTS

Many of the technical developments that made radio possible went on to become the foundation for many other applications such as TV, Microwave, Satellite, and Cellular Applications.

A. ELECTROMAGNETIC RADIATION EXPLAINED

1. Maxwell's Equations--a Brief Explanation

The most significant mathematical foundations for electromagnetic theories were provided by Maxwell when, in 1873, he published a series of differential equations (Maxwell's Equations). His equations described the relationships between electricity and magnetism and can be written as:

$$\nabla \times E = -M_i - j\omega\mu H$$

$$\nabla \times H = -J_i - j\omega\varepsilon$$

$$\nabla \bullet (\varepsilon E) = qve$$

$$\nabla \bullet (\mu H) = qvm$$

Where ∇represents...

Ji represents *electric current* density,

Mi represents *magnetic current* density,

qve represents *electric charge* density,

qvm represents *magnetic charge* density,

E represents the *intensity of the electric field,*

H represents the *intensity of the magnetic field,*

μ the *voltage amplification factor*, ε the *permittivity,*

and ω represents *2π x frequency.*

2. Electromagnetic Wave Components

As one might speculate after reviewing the famous equations of Maxwell, electromagnetic waves are comprised of two primary fields--electric and magnetic. An *electric field* is created by the force of voltage (the relative imbalance of electrons). For example, when one terminal of a battery is positively charged, it is said to have a relative deficiency of electrons. Conversely, when a terminal of a battery possesses a relative surplus of electrons it is said to be negatively charged. This relative difference, or imbalance, of electrons comprises the voltage of the battery. When a conductor of electrons is connected to both terminals of the battery, surplus electrons flow from the negative terminal to the positive terminal. This flow of electrons is also known as an electric current. Further, the flow of electric current produces a *magnetic field.*

An electric field is simply a line of force (capable of causing electric charges to move within it) around an electrically charged object. Such lines of force are typically shown in diagrams as the path along which a single positive charge would move when acted upon by the field. One interesting aspect of such lines of force is that they repel each other when several are present and acting in the same direction. Electric fields (also known as E Fields) are illustrated in Figure 1 . Note that the devices shown below are also known as capacitors. A capacitor is simply a device capable of storing electric energy in a *field.*

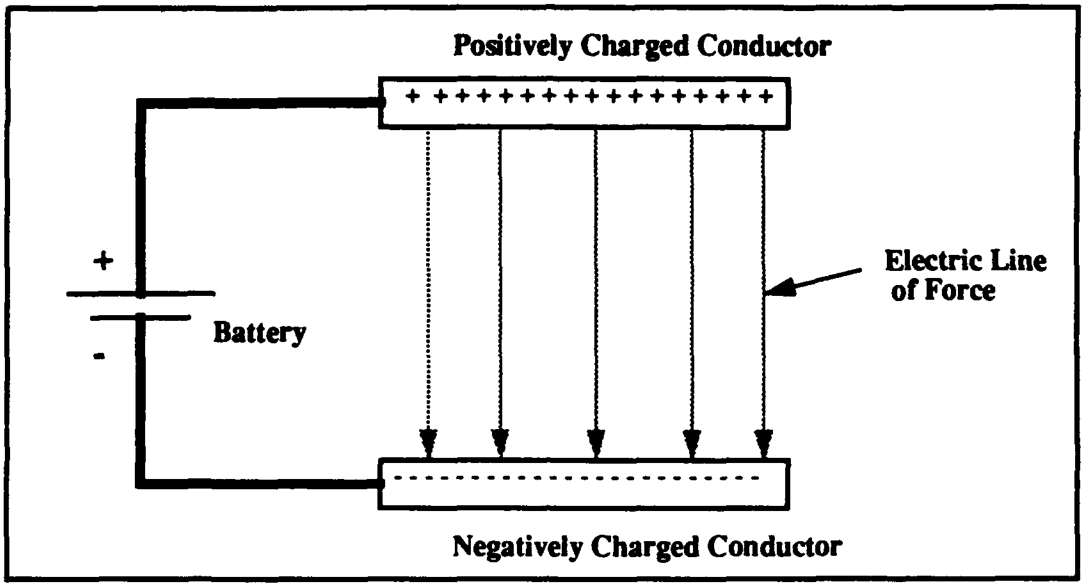


FIGURE 1. A CAPACITOR

Magnetic fields are created in the area *surrounding* the conductor of any electric current. Unlike electric fields, magnetic fields that are present in parallel and acting in the same direction reinforce each other--creating and even stronger magnetic field. Figure 2 illustrates the direction of magnetic lines of force (also known as H Lines).

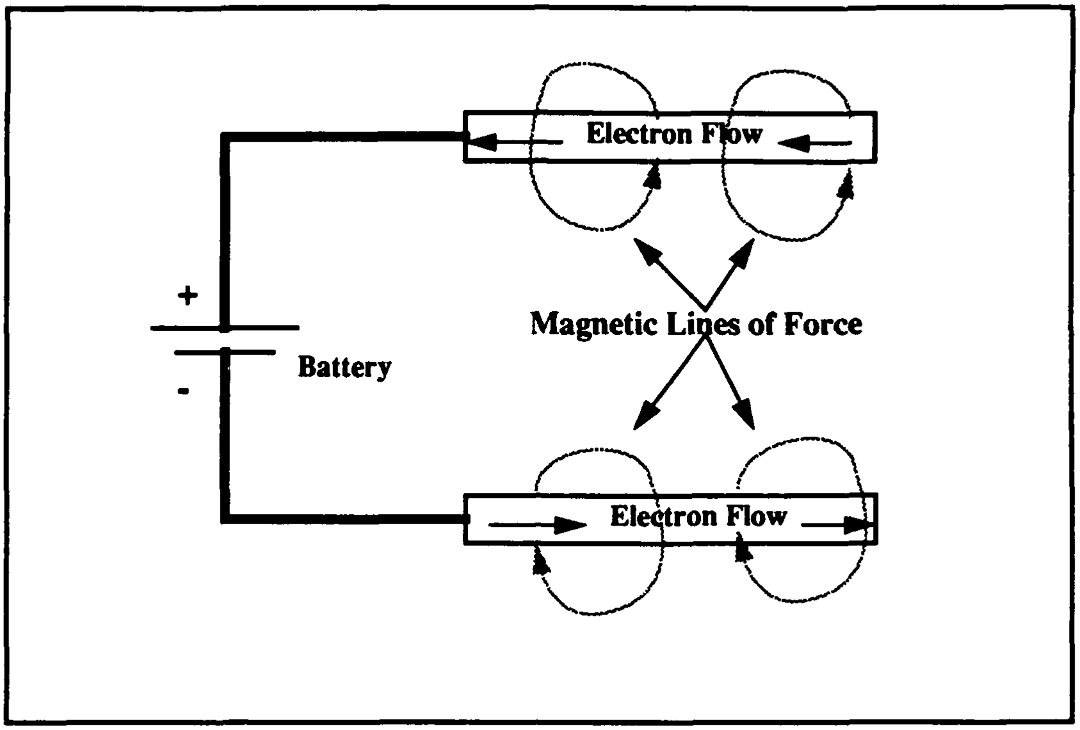


FIGURE 2. MAGNETIC LINES OF FORCE

3. Radiation of Electromagnetic Fields

The process of radiating (or broadcasting) electromagnetic fields can be shown if the capacitor diagrams shown above are modified such that the conductors are extended away from each other (rather than parallel). Such a configuration is commonly referred to as a half-wave dipole antenna. Notice in Figure 3 that the new configuration allows the magnetic fields to reinforce each other since the current is flowing in the same direction through each conductor.

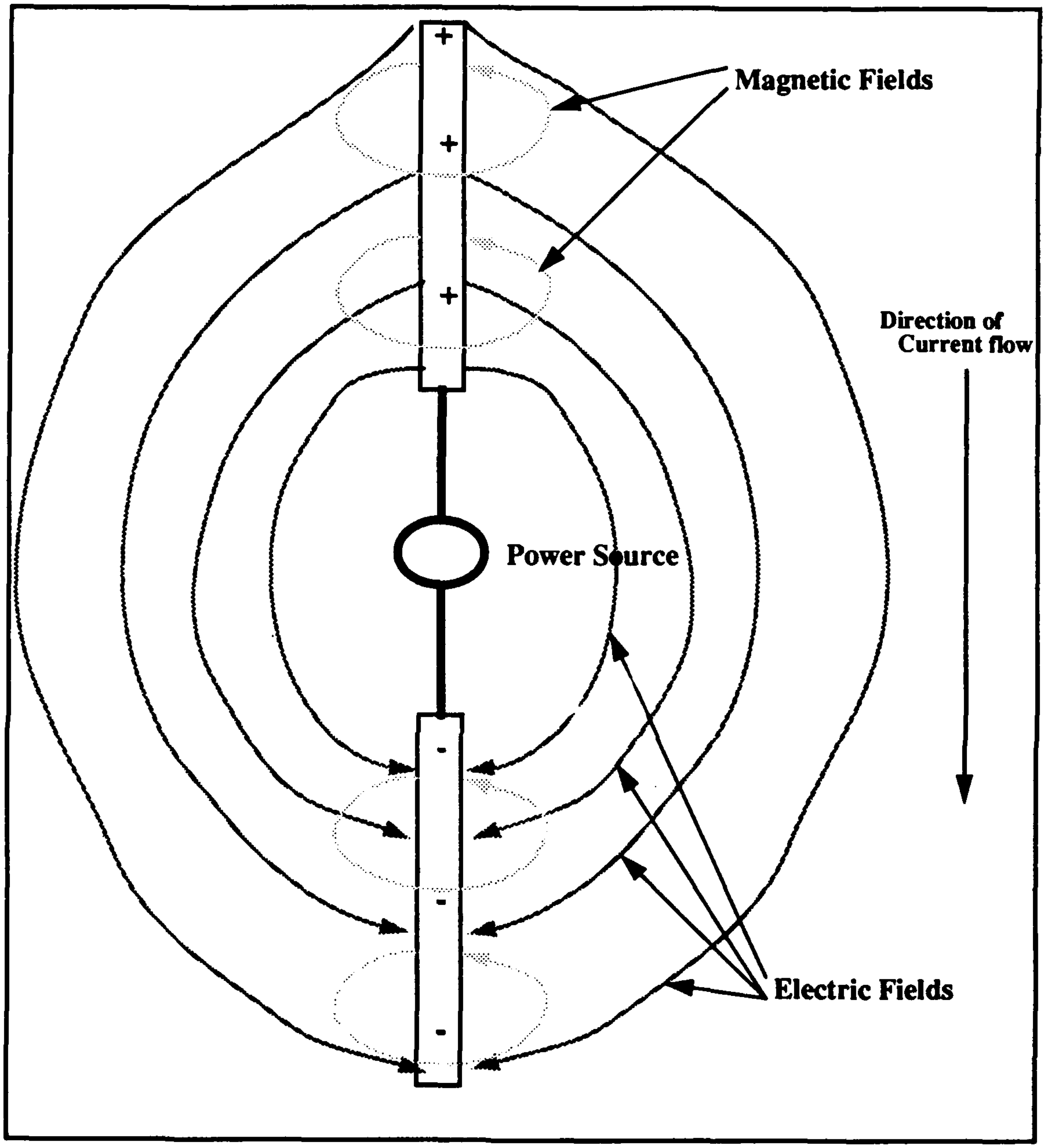


FIGURE 3. HALF-WAVE DIPOLE ANTENNA

Using such an antenna with a power source that alternates its current allows for the expulsion of electromagnetic forces from it. It too can be illustrated by the following diagrams. As maximum voltage is applied by the power source in one direction, the first antenna is shown where its electric lines of force begin and end at the ends of the antennas (the magnetic lines of force are not shown). As voltage drops, the second antenna depicts the lines of electric force beginning and ending near the centers of each antenna.

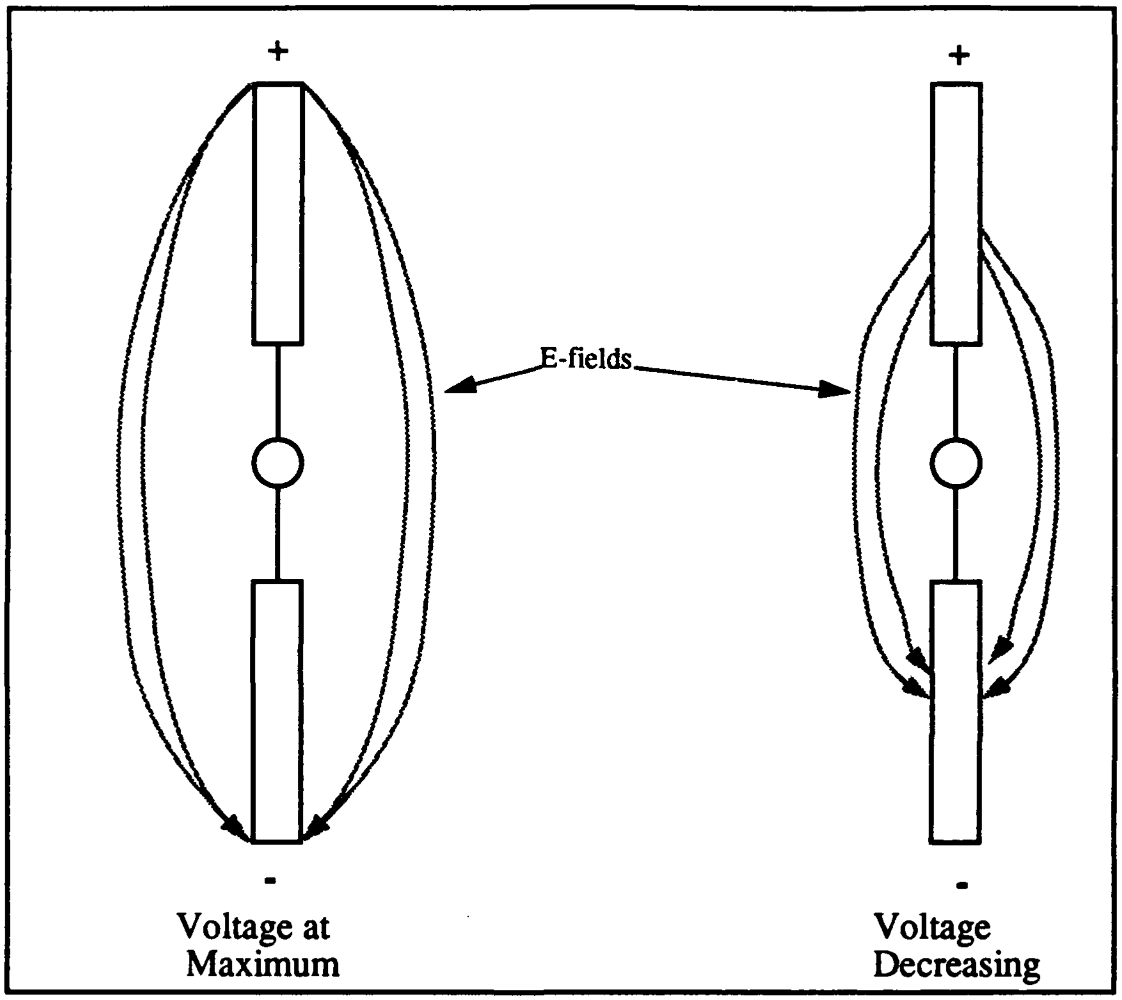


FIGURE 4. E-FIELDS AT MAXIMUM & DECLINING VOLTAGE

As the voltage drops to zero, as it must prior to changing direction of the current, some of the lines of electric force come together to form complete loops (detaching themselves from the antenna completely). Finally, as the current direction is reversed and the voltage increases, electric lines of force are again generated which begin and end near the centers of the antennas. However, with the direction of the current reversed, the direction of the electric lines of force is also reversed. Since the direction of the electric lines of force coming from the antenna is the same as those detached from the antenna, the detached loops are repelled away from the antenna at the speed of light[DAFM85].

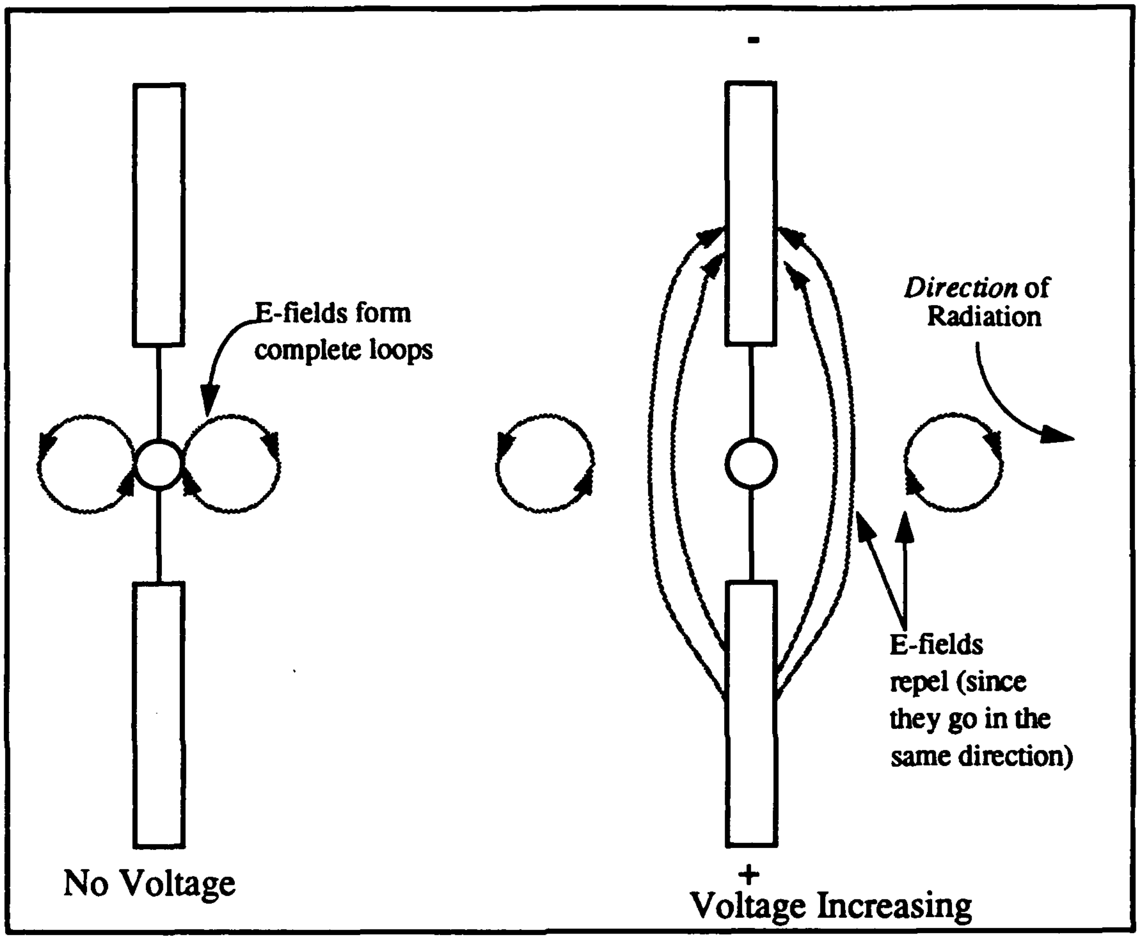


FIGURE 5. E-FIELDS DURING PERIODS OF NO VOLTAGE AND INCREASING VOLTAGE

Note too, that the direction of radiation is perpendicular to the antenna being used. While relatively minor amounts of electromagnetic radiation are sent off in other directions, the primary vectors along which the majority of the radiation travels away from the antenna are perpendicular. The intensity of the radiation is a function of the angle of departure from the antenna--maximal at right angles and non-existent at the ends of the antenna. Some antenna designs are more directional than others. The ratio of the amount of energy focused in certain directions to the amount of energy that would be propagated in a non-directional antenna is known as the *gain* of the antenna. This directional nature of wave propagation is quite important in orienting both the transmitting and receiving antennas.

B. RADIO COMPONENTS & HOW THEY WORKED

Many components have been developed to improve upon the radio and subsequent wireless applications. A few of the more basic ones are described in more detail below.

1. Spark Transmitters

The very first radios developed did not modulate a voice over a continuous wave for transmission. Rather, quite like the telegraph, they transmitted information in digital form as dots and dashes through the air via use of the "spark gap" transmitter developed by Hertz. This relatively simple arrangement consisted of a power supply and a single electric circuit.

a. Power Supply

Although power supplies have improved with time, the earliest types of power supplies stored static electricity in simple devices known as Leyden Jars. The late 18th century saw the development of the first batteries or voltaic cells, which were developed by Volta. Such "batteries" were used by Hertz in his experiments with spark transmitters to provide a direct current to his radio circuit.

b. Basic Circuit

A simple circuit comprised the first Hertzian transmitter. The power supply was connected to a circuit that contained a very small gap of air between the two ends of the conductors that comprised the circuit. When the difference in charges on both sides of this small gap became great enough (due to a build up of electrons on the side connected to the negative terminal of the battery), the electrons would use the air as a conductor in order to jump or "spark" across the gap in the circuit.

As the current flowed across the gap, an excess of electrons would cross over the positive end of the circuit because of the momentum of the current. This resulted in the need to cross back across the gap to the original side in order to bring the relative charges of both sides closer into balance. The process repeated over and over until a relative balance of charges was achieved on both sides of the gap. A visible spark was actually caused by this back-and-forth oscillation of the current. Further, in accordance with Maxwell's equations, the *acceleration* of the oscillatory current generated electromagnetic waves. The *frequency* of the transmission could be adjusted by moving the metal plates or metal spheres at each end of the dipole antenna used in the circuit. (This changed the frequency of oscillation.) The waves could be detected at a short distance from where the sparks were actually generated. Thus, by connecting the power supply to such a circuit only when one wanted to generate such a wave, a signal could be sent. This was easily accomplished by connecting a telegraph key to the circuit.

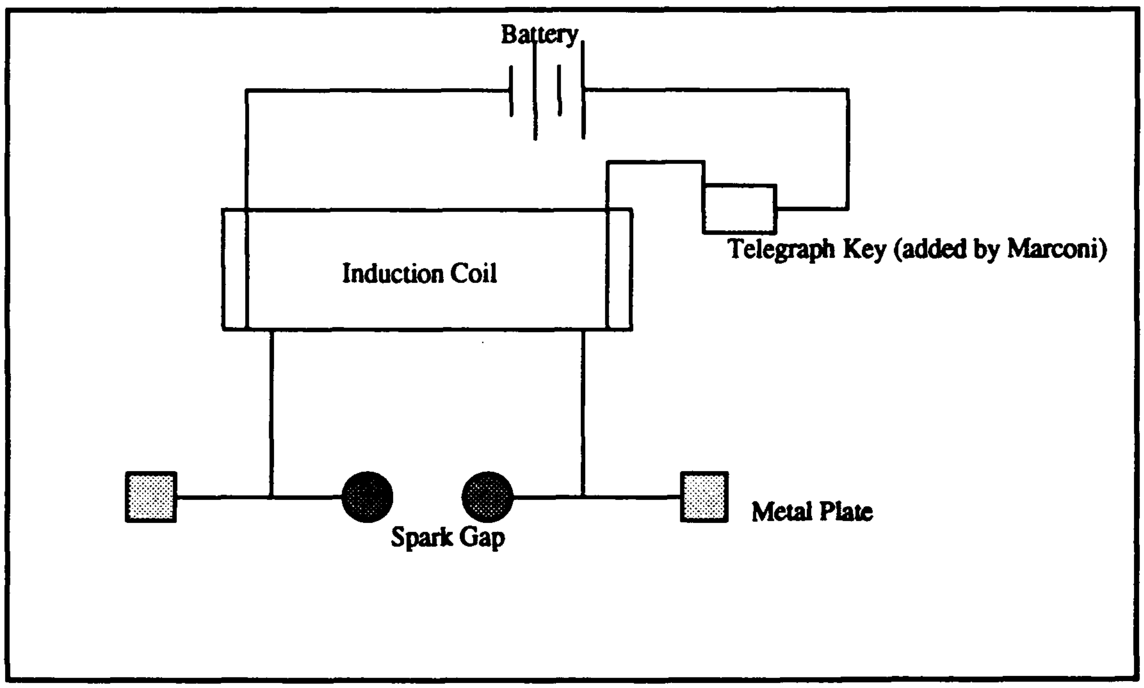


FIGURE 6. THE HERTZ OSCILLATOR

2. Continuous Wave Transmitters

In order to transmit voices, music, and other non-digital signals, a transmission method more sophisticated than a spark transmitter was needed. Spark transmitters employed a direct current (DC power) to turn on or off a signal's transmission by completing the circuit with the tap of a telegraph key. Voices, music, and other analog data are continuous in nature rather than discrete. Accordingly, the continuous sound waves that comprise them must be superimposed onto radio waves. This is accomplished via a modulator--a device which transforms high-frequency radio waves of uniform amplitude into waves whose amplitude matches that of the underlying audio frequency to be sent. These resulting waves are used to control the output of the transmitting oscillator that, in turn, creates electromagnetic waves. The oscillator is connected to an aerial (or antenna--one or more wires sticking up into the air) that has high-frequency alternating current (AC) flowing through it. The result is the transmission of the electromagnetic waves from the oscillator into space.

3. Coherers and Receivers

Obviously, in order for a radio to be of any value, transmissions must be received. The very first Hertzian waves were detected at only a few feet from the source of transmission. In order to detect transmissions it was necessary to construct a receiving loop of wire that was identical in resonance to the transmitting circuit. In other words it had to be designed to detect only waves of a specific frequency, that of the transmitter's. This is similar to the observation that two (separated) tuning forks in the same room will both vibrate when one is hit only if they are identically designed to emit the same pitch or frequency. In the experiments of Hertz, the spark created in his circular resonator were so minute that he had to darken the lab and use a magnifying glass to detect the presence of a spark in the receiving loop. For detection at greater distances, other components needed to be developed.

a. *Early (Spark) Receivers*

As described earlier, Branley discovered a way to detect waves generated by spark transmitters using a device known as the coherer. Many varieties of the coherer were developed but they all employed the same basic principles to detect the waves. The coherer consisted of a transparent tube (usually glass) that was partially filled with electrically conductive metal filings. The tube was usually evacuated as best as possible, then plugged on each end with a (conducting) metal plug. This device was inserted into a receiving circuit. The presence of electromagnetic waves caused the metal filings in the coherer to "cohere" or stick together allowing a current to flow from one end of the tube to the other. This was not too unlike metal filings being caused to move into specific patterns with respect to a nearby magnet. The cohered particles, once connected, had to be shaken up or tapped apart in order to detect subsequent signals. Thus, a companion device to the coherer, known as a tapper, was also quickly developed. The earliest coherers detected spark-gap transmissions up to 25 yards away. As the coherer was improved upon with various other metal filing combinations and better vacuums this distance was increased considerably. The coherer was often used in combination with some sort of antenna or aerial to further increase its sensitivity.

The early receiving circuit was simply composed of a power source (battery) with each of its terminals connected to one end of the coherer. The coherence of the metal filings was visibly noticeable when electromagnetic waves caused them to cohere. The addition of an audible signal to the circuit, such as a bell or loud speaker, created noise whenever electromagnetic waves were detected. While numerous improvements were made to the power supplies, transmitters, and aerials of spark-type radio systems, the coherer changed very little during its thirty years of use.

b. *Voice Receivers*

Once voices were transmitted using continuous wave transmissions, devices such as the coherer were useless for such receptions. Instead, the high-frequency electromagnetic waves must be received (usually by aerial) and demodulated to separate the original audible signals from the high-frequency radio wave (carrier wave). The resulting signal is then amplified and sent through a loud speaker to reproduce the originally transmitted sounds.

4. Aerials

For detection of signals at ever greater distances, another component needed to be developed. It was quickly discovered that by adding an antenna to the transmitter and/or receiver, the effective range of the transmission could be greatly increased. Thus aerials (or antennas) soon became one of the most critical components of radios.

a. Early Antennas

From the earliest uses of antennas until about World War II, the primary technology of antennas focused on the use of one or more wires suspended in the air.[BALA 92]. The earliest antennas, known today as long-wire antennas, were simple to construct and worked well over a range of frequencies. Such antennas are one or more wavelengths in length with respect to the frequency at which they will operate. Unfortunately, their gain (directional characteristics of their radiation) is lower than more modern antennas. They do however, have the added advantage of being able to radiate fairly efficiently at any operating frequency for which the long-wire antenna's length is greater than approximately one-half wavelength. For these reasons, they were among the first generally used antennas and can still be found in use today.

In the transmission of radio waves, the antenna is attached to the transmitting circuit as described in Section 2. above. As alternating current is pumped into the antenna, electrons flow to the end of the antenna and return to their source. As Maxwell predicted, the acceleration of the electrons in such an oscillatory manner generates electromagnetic waves that, in the case of a straight-wire antenna, emanate in all directions. In the case of other than long-wire antennas, it becomes critical that the length of the transmitting antenna be such that the electrons returning from the far end of the antenna arrive as the next cycle of the radio frequency voltage begins. This can be accomplished by using an antenna that is one half the length of (one cycle of) the wave being generated. The round-trip takes exactly one cycle under such circumstances. Such antennas are known as Hz (or half-wave) antennas. They are most practical for radio frequencies above 2 MHz.

An alternating current feeding into one end of an antenna ideally alternates at a frequency that accomplishes the above phenomenon. As waves travel to the far end of the antenna and back, they are reinforced by the next cycle of the current feeding into it. When the frequency of the current (or alternatively, the length of the antenna) does not allow this to happen, the waves tend to cancel each other out and lose their energy in the form of heat. However, when ideal conditions are achieved, the reinforcing of each subsequent wave reflection creates what is known as resonance or synchronization within the antenna circuit.

On the receiving end, the electromagnetic field impacting upon the receiving antenna causes a voltage to be induced within the antenna. In theory, the efficiency with which an (identical) antenna converts the electromagnetic waves into electrical voltages (at the receiving end) is the same as the efficiency for the reverse process (at the broadcasting end).

For frequencies (primarily) below 2 MHz, another commonly used antenna is the Marconi (or quarter-wave) antenna. Such antennas are widely used in the military as well as for mobile transmission and reception.[DAFM85] As its name suggests, a quarter-wave antenna is only half the length of a half-wave antenna. Such an antenna, *when grounded*, will resonate at a frequency that corresponds to a wavelength 4 times the length of the antenna. The fact that the antenna is grounded, unlike the half-wavelength versions, allows its length to be only half as long. The reason for this is that the ground "acts as an electrical mirror image"[DAFM85], providing the "missing half" of the antenna as depicted in the following diagram.

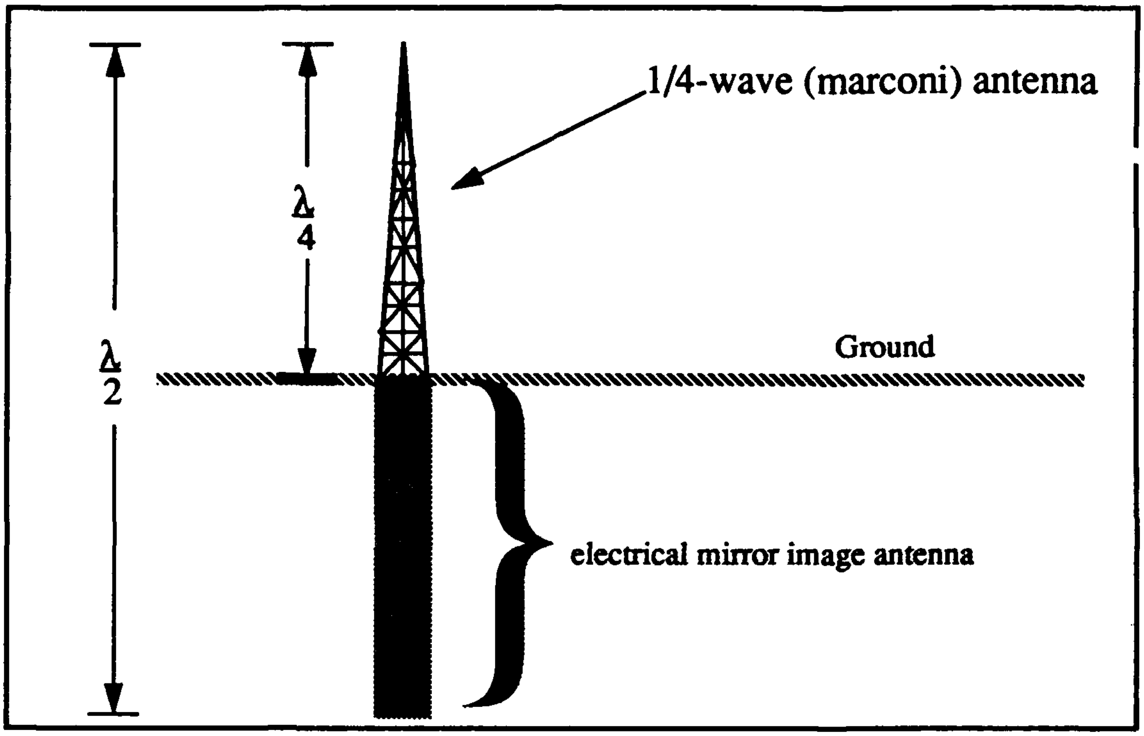


FIGURE 7. QUARTER-WAVE ANTENNA

b. Modern Antennas

Since the mid 1940's, antenna technology has rapidly grown. Today a wide variety of antennas have been developed for a wide variety of applications. Modern computers make possible the optimization of the complex equations that dictate antenna performance. Accordingly, antennas have become more specialized, more powerful, and smaller. They can be made highly directional or very sensitive to specific frequencies. Antenna evolution today is still continuing at a rapid pace.

5. Tuners

With several radios in operation simultaneously in a given geographical area, it is usually the goal of the receiver of such signals to focus attention on a single broadcast. Such focus requires an ability to tune or adjust the frequency of transmission and reception so that they are both the same. The very first Hertzian waves were detected at only a few meters from the source of transmission. Hertz himself was able to do this. For detection at greater distances, other components needed to be developed. The ability to tune transmitters and receivers was critical to the development of the radio. If everyone used the same frequency for radio transmissions, there would be so much interference from other users that no one would be able to successfully receive any transmission with accuracy.

a. Early Tuners

Oliver Lodge developed a method of tuning transmitting and receiving circuits that he patented in 1897.[AITK76] His central idea was to simply to adjust both the transmitting and receiving circuits in such a way that they were capable of oscillating at a particular frequency and detecting such frequencies respectively. Specifically, he focused on tuning the antennas of both receiver and sending unit. This was most easily accomplished by varying the length of the antennas involved.

b. Modern Tuners

When it is not practical to physically lengthen or shorten the antenna in order to bring it into resonance with the frequency at which one is trying to operate, the *electric* length of the antenna can be increased or decreased by increasing the *inductive* or *capacitive* reactance respectively. This method of antenna tuning is known as antenna loading.

6. Amplifiers

Due to the fact that waves diminish in amplitude as they travel through space (attenuation), the need to amplify received signals rapidly became apparent to those involved in radio signaling. The problem was eventually solved by Lee DeForest in 1906 with the invention of what he called the Audion.

a. Edison's Two-Electrode Tube

The origins of the means of modern amplifiers can be traced back to the great inventor, Thomas Edison. In the 1880's, Edison attempted to improve upon designs for his incandescent light by adding a new element to the bulb--a metal plate. To review, the incandescent bulb operates by passing a current through a fine filament in a vacuum. The flow of electrons causes the filament to glow as its temperature increases (known also as a thermionic emission--the escaping of free electrons from the filament). The vacuum prevents the filament from burning.

The effect of adding a (positively-charged) metal plate to the bulb was that Edison was able to detect the passage of a current across the vacuum from the filament to the vacuum. The flow of electrons across this gap came to be known as the *Edison Effect.*

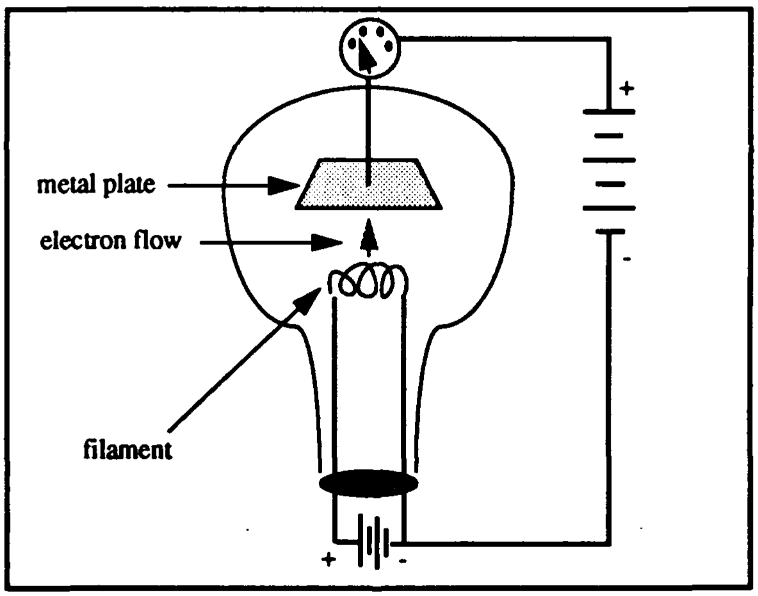


FIGURE 8. EDISON'S MODIFIED TWO-ELECTRODE TUBE (THE FLEMING VALVE)

b. The Fleming Valve (Diode)

The two-electrode tube invented by Edison was modified by Dr. J. Ambrose Fleming in 1905. Fleming reversed the charge on the metal plate (making it relatively negative instead of positive) with respect to the negatively charged filament. He was able to determine that the flow of electrons from the filament to the plate was shut off--like a valve. Most significantly, Fleming attached an alternating current to the plate during his experiments. The result was that electrons only flowed during the positive half cycle of the alternating current--effectively converting the alternating current to a direct current (albeit a pulsating one). This device became the first of many *vacuum tubes* used extensively in electronics for decades to follow.

c. *The Audion (Triode)*

Lee DeForest modified Thomas Edison's two-electrode tube in 1906 when he developed the Audion that allowed voice amplification.[ARCH38]. Until that point in time, the attenuation of both radio and phone signals had been a severely limiting factor for both means of communication over great distances. What was needed was a means of amplifying signals to combat the effects of attenuation.

During the process of experimentation, DeForest added a tiny wire grid to the Fleming Valve. He inserted it between the electron emitting filament and the metal plate and attached it to a relatively weak power source. DeForest noted that a small amount of positive voltage applied to the wire grid dramatically increased the flow of electrons from the filament to the metal plate. Conversely, a small negative voltage applied to the wire grid caused a dramatic reduction in the rate at which electrons migrated from the filament to the metal plate. This happened for two reasons. First, the distance between the filament and wire grid was significantly less than the distance between filament and metal plate. Thus, the attraction of the relatively weak positive charge of the grid was enhanced considerably by its proximity to the filament. Second, the wire grid was a very loose mesh of very fine wire. Accordingly, the vast majority of the electrons that were migrating from the filament toward the grid simply passed on through the holes in the grid and were subsequently attracted to the metal plate.

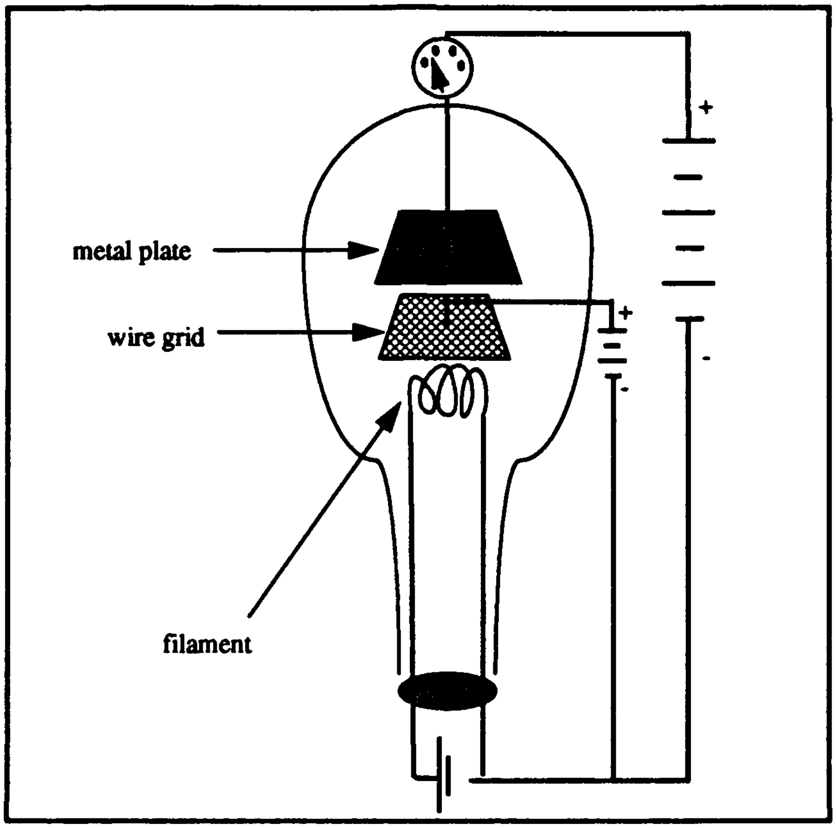


FIGURE 9. DE FOREST'S AUDION

While the amplifying tube known as the triode (or Audion) became critical in the advancement of electronics in general, DeForest applied its power to voice amplification in radio transmission specifically. By attaching the aerial of the receiver to the wire grid of the Audion, the incoming electro-magnetic waves would boost or decrease the original flow of electrons across the main circuit--effectively amplifying the incoming signal. In addition, several Audions could be used in succession to increase the amplification to any desired level. This made large, remotely-located antennas quickly obsolete for the reception of trans-Atlantic and other distant radio transmissions.

7. Modulators

The function of a modulator is to combine the audible analog sound waves to be transmitted with the inaudible high-frequency radio waves that will carry them over greater distances. The process must be done in such a manner that it is reversible at the receiving end (i.e. demodulated or separated).

a. Early Modulators

Originally, the first modulators developed utilized the electrical output of a microphone as input into the control grid of a triode (or Audion) within the radio circuit. At the same time, the radio frequency alternating current was fed through the triode. This had the effect of increasing the radio frequency's power in synchronization with the power of the microphone output. The end result was the modulation of the audible signals to be sent onto a radio frequency alternating current.

b. Modern Modulators

More complicated methods are used today that overcome the problems of frequency and power distortion which were the drawbacks of the earlier simpler methods of modulation. For the most part, modern methods result in improved modulation quality at the cost of considerably higher power requirements and more complicated equipment. Research continues today on improving methods of modulation.

8. RF Generators

Central to transmission of analog signals over radio waves was the development of a radio frequency generator upon which the audible analog signals could be modulated prior to transmission. Such generators must be capable of generating a constant and consistent high-frequency in order to be detectable by a tuned receiving station. Generating such high frequency waves with accuracy was a major stumbling block in early voice transmission over radio waves. The need for precise generation of radio frequency power was critical in order to perfect the tuning of the radio circuit and keep various broadcasts within their own "channel" or precise radio frequency. Otherwise, broadcasters on nearby (neighboring) frequencies would experience interference from adjacent frequency broadcasts.

a. Early Alternators

Alternating current generators are known as alternators. The earliest alternators were turbine-type generators that produced the current via rotating magnets. The major drawbacks to this method were the limited, relatively low frequencies that could be produced. Since the frequency produced was effectively a function of the rate of rotation of the magnets, high frequency alternators needed very high rates of rotation. Additionally, the power output of the early generators was often quite limited.

Alternators capable of 10,000 cycles per second were relatively easily built as early as 1903. Unfortunately, this relatively low frequency was unsatisfactory for radio transmission of the entire audible range. One of the most famous of the early alternators was built by Ernst F.W. Alexanderson of the General Electric Company in 1909 that was capable of producing 2,000 watts of power at 100,000 cycles per second[ARCH38]. By 1918, another Alexanderson alternator, capable of 200,000 watts was used by President Wilson during World War I to address the German people directly from a station in New Jersey.[ARCH38]

b. Modern R.F. Generators

Today's radio frequencies are generated by electronic components known as oscillators. The oscillators can be configured to generate highly precise and uniform high frequency current. Their power output is, however, quite low and must be increased via high-efficiency amplifiers prior to being useful.

9. Loudspeakers

Loud speakers can be thought of as the opposite of microphones. Whereas the microphone can be regarded as an audio-frequency generator translating audio-frequency air waves into electromagnetic waves of the same frequency, the loud speaker must translate in the other direction. This transition necessarily utilizes the electromagnetic received to vibrate or stir up the air waves in order to reproduce a sound similar to the one originally transmitted. Several types of loudspeakers have been developed, but the one in common use today is the moving-coil type.

a. Moving-coil Loudspeaker

The interaction of two magnetic fields is used to move a conical diaphragm in this type of speaker. One magnetic field is held constant, while the other is controlled and varied by the passing signal currents through a movable wire coil. The field held constant is created by a ring shaped circular magnet. Its strong magnetic current creates an intense magnetic flux across a narrow circular gap that exists near the center of the conical diaphragm. A *moving-coil* is suspended in this gap. When the alternating signal currents are passed through the moving coil, the coil moves back and forth in the gap. Since the moving coil is attached to the center of the diaphragm, the diaphragm vibrates in tandem with the coil. This vibration stirs up the air in a manner which produces the audible signals desired.[SCRO71]

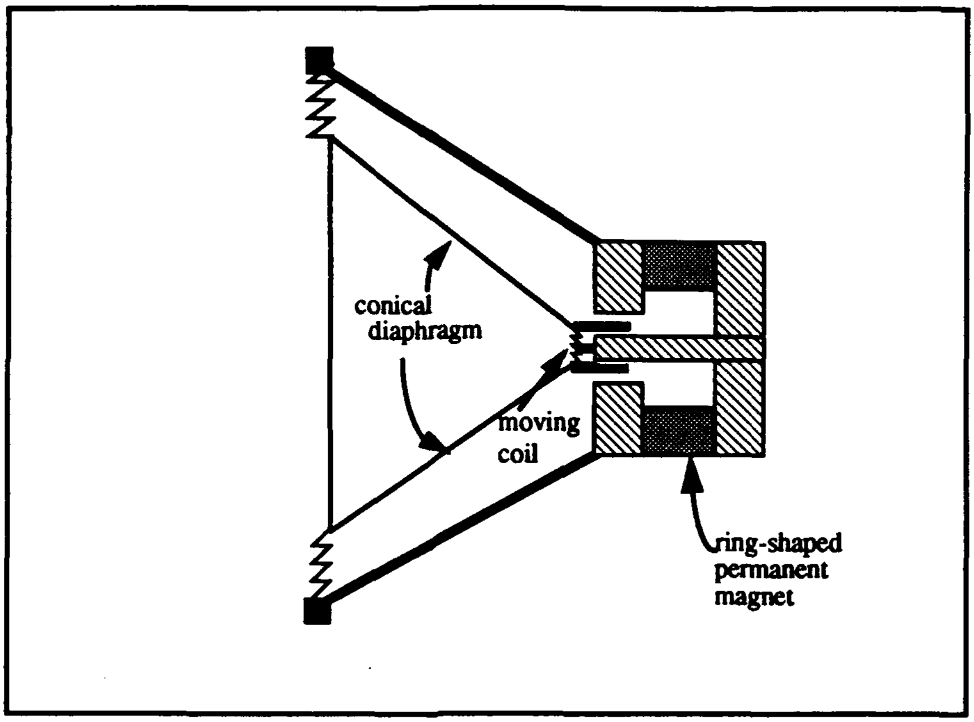


FIGURE 10. CROSS-SECTION OF A MOVING-COIL LOUDSPEAKER

b. Improvements on the Loudspeaker

Efforts to improve sound quality have led to the use of separate loudspeakers for use in each region of audible frequencies. For example, the reproduction of very low frequency sounds requires that large volumes of air be moved. This can be accomplished by either increasing the amplitude movement of the moving coil or using a larger cone. The use of larger movements by the magnetic coil would require a larger, more powerful magnet. Alternatively, if a larger cone is used it becomes difficult to get the cone to vibrate as a whole at the higher frequencies and distortion occurs. The solution used today is simply to use a series of loudspeakers for the various audible frequency ranges. Large speaker cones are used for the lower "base-type" tones while smaller and smaller cones are used for the higher-pitched tones.

10. Transistors

Electronics entered a new era with the development of the transistor in late 1947. The bulk of circuitry in all modern radios is comprised of transistors. Transistors have increased the reliability while decreasing the size and cost of virtually all electronic components from radios and televisions to computers and medical equipment. Transistors are devices that allow the flow of current through them to be controlled by a third electrical connection--in many respects similar to the triode. Unlike its vacuum tube ancestor, transistors are (at a minimum) thousands of times smaller than triodes. There are two types of transistors, *bipolar* and *field-effect.*

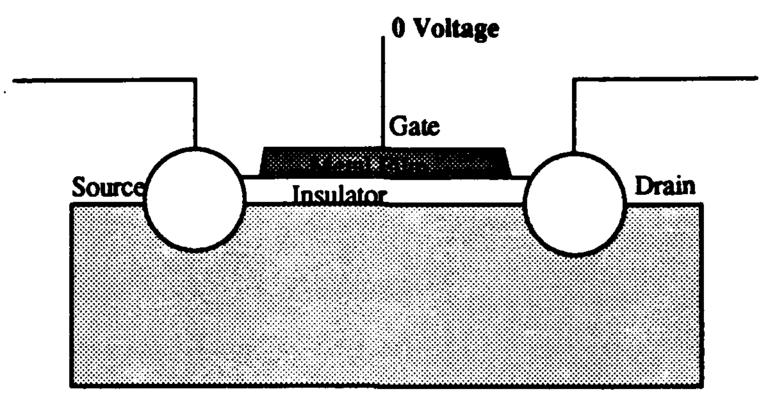
a. Bipolar Transistors

The faster switching (and more difficult to construct) type of transistor is known as the bipolar transistor. Charges are carried in a bipolar transistor using both electrons and holes (positive charge carriers). The three main parts to a bipolar transistor are the emitter, base, and collector.

Charged particles flow primarily from the base to the collector. When the emitter injects charged particles into the base, the flow of current from the base to the collector stops. When no particles are being injected, the flow continues normally.

b. Field-effect Transistors (FETs)

Unlike Bipolar transistors, FETs use only electrons *or* holes as charge carriers. Their main regions are the source, drain, gate, and channel. The source and drain are simply the segment of the circuit through which one wants to control the flow of current via a third electrical connection (the gate). When no current is applied to the gate, the insulator between the source and drain prevents current from flowing from source to drain.



Field-effect transistor without voltage
applied to its third electrical connection

FIGURE 11. FIELD-EFFECT TRANSISTOR--NO VOLTAGE

When positive voltage is applied to the metal film above the insulator (the gate), electrons are attracted from the material below the insulator. These electrons change the conductive property of the substrate or base material upon which the transistor is fabricated in such a manner as to form a channel through which current can flow from source to drain.

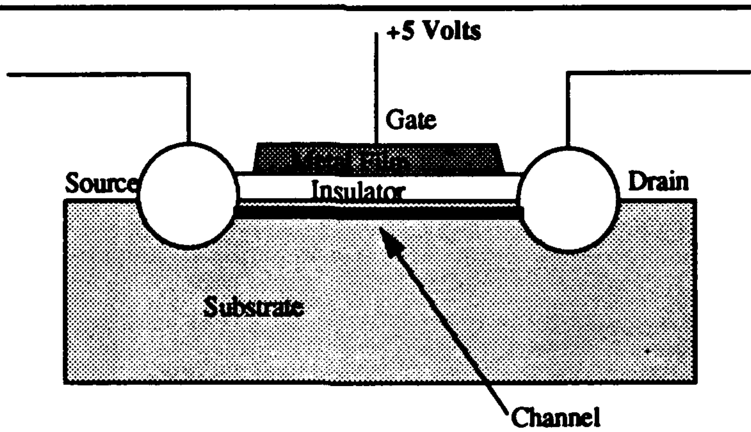


FIGURE 12. FIELD-EFFECT TRANSISTOR WITH VOLTAGE APPLIED

11. Integrated Circuits

In 1958, the first integrated circuit (IC) was developed by Jack Kilby of Texas Instruments. It consisted of "two circuits placed on a single piece of germanium."[RYAN90] Integrated circuits of today have been scaled down to the point where more than a million transistors per IC chip can be commercially mass produced.

a. Bipolar-junction-transistor-based IC

As the name implies, bipolar-junction-transistor-based integrated circuits are comprised of multiple bipolar-junction transistors. While they have the advantage of very fast switching times, this type of integrated circuit generates considerable heat during operation. The heat generated is the primary limiting factor in the determination of how many transistors can be placed on a single IC. Fewer than 100,000 bipolar-junction transistors can be successfully placed on commercial ICs[RYAN90].

b. Metal-oxide Semiconductor (MOS) Transistor-based IC

Very Large Scale Integration (VLSI) of circuits (more than 100,000) was made possible only by utilizing the cooler, slower metal-oxide semiconductor (MOS) technology with field-effect transistors. MOS technology simply uses the *oxide* of the substrate (base) material as the insulator in the FET. A substrate of silicon would, for example, call for an insulating layer of silicon *dioxide*.

C. THE ELECTROMAGNETIC SPECTRUM

The frequencies of electro-magnetic energy vary considerably. Electro-magnetic energy with relatively low frequency in known a radio waves. Higher in the spectrum we find Infrared Rays, visible light, X-Rays, Gamma Rays, and Cosmic Rays. Each band of the spectru· ·~s uses (or potential for use) in the field of communications. The table below summarizes the current communication-related uses for each band [FREE91].

TABLE 2: THE ELECTROMAGNETIC SPECTRUM

Frequency	Wavelength	Name	Applications/Uses
< 3 KHz	> 10Km	ELF	Submarine Communication
3 - 30 KHz	10-100Km	VLF	Military / Navigation
30 - 300 KHz	1 -10Km	LF	Medium-range Navigation
300 KHz - 3 MHz	100-1000m	MF	AM radio / Ship-to-Shore
3 - 30 MHz	10 - 100m	HF	Shortwave Broadcasting
30 - 300 MHz	1 - 10m	VHF	TV, FM, Cellular Phones
300 MHz - 3 GHz	10cm - 1m	UHF	TV, Land-mobile, Radar
3 - 30 GHz	1 - 10cm	SHF	Microwave, Terrestrial & Satellite Relay
30 - 300 GHz	1mm - 1cm	EHF	Space-to-Space Satellite Links
10^3 - 10^7 GHz	< 1/3mm	light	Fiber Optic transmission

Commercialization of the radio quickly brought about the need for regulation of the use of the electro-magnetic spectrum (or "air waves"). Today, for example, AM radio uses the part of the spectrum between 535 KHz and 1,605 KHz. FM radio broadcasts to the general public uses the band between 88 and 108 mHz. Each station uses a channel (or slice of bandwidth) 200 kHz wide for FM (10 kHz for AM). Such channel allocations allow for 107 frequency locations for AM station broadcasts. Limitations on the power of broadcasts further restrict each broadcaster to a limited geographic region of the country. Television uses part of the spectrum above and below that of FM radio.

D. AMPLITUDE MODULATION

Amplitude modulation or AM radio was the first type of radio transmission to carry analog signals. It operated by modulating an audible voice frequency onto a much higher radio frequency prior to transmission. The (radio) frequency was kept constant while the amplitude of the radio wave was varied in such a way that the difference in amplitude over time paralleled the audible sound wave to be transmitted. This can be illustrated in the three figures that follow. The wave shown in Figure 13 is the audible sound wave that we wish to transmit. Its frequency, in cycles per second, varies within the audible range of 20-20,000 Hz. The second wave, shown in Figure 14, is a radio-frequency wave generated by the radio-frequency generator. It is simply a sine wave with a relatively high radio frequency--it carries no information by itself. The third diagram, Figure 15, illustrates how the amplitude of the radio frequency wave can be adjusted in order to carry the information of the original sound wave we wish to transmit. This adjusted radio-frequency wave (bold wave in Figure 15) is the one we will transmit. By connecting the peaks and valleys of these waves (the non-bold wave in the third diagram), we can recreate our original sound wave at the receiving station.

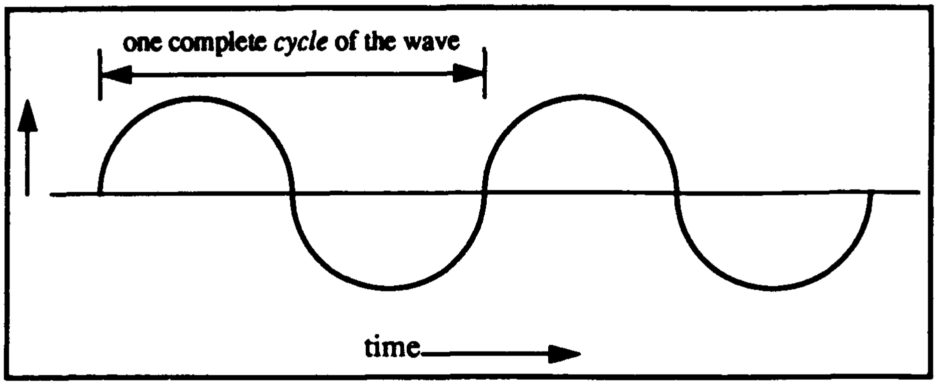


FIGURE 13. ORIGINAL LOW-FREQUENCY AUDIBLE SOUND WAVE TO BE BROADCAST

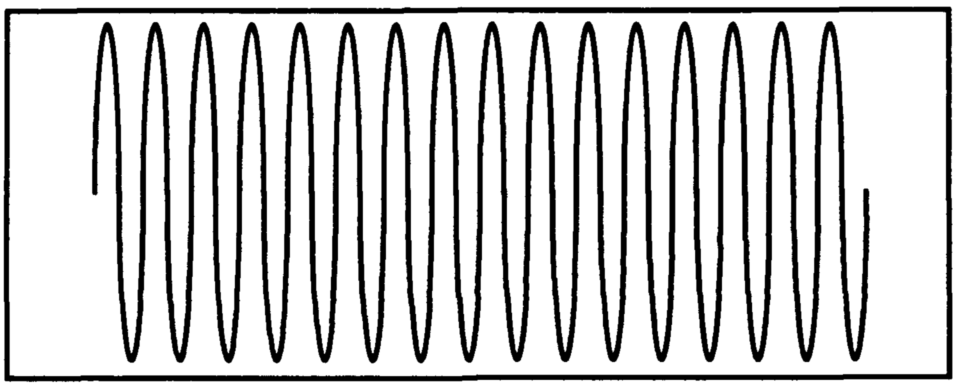


FIGURE 14. RADIO-FREQUENCY CARRIER WAVE GENERATED FOR BROADCAST OF THE SOUND

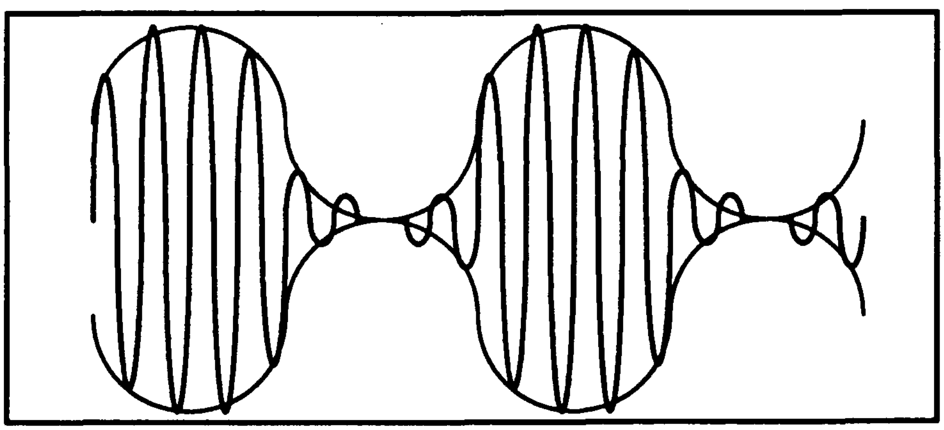


FIGURE 15. CARRIER WAVE WITH AMPLITUDE ADJUSTED TO "CARRY" THE SOUND WAVE

E. FREQUENCY MODULATION

Instead of varying the amplitude of the radio frequency wave while keeping the frequency constant, FM (frequency modulation) does just the reverse. The amplitude of the radio wave is kept constant while increases and decreases in the frequency of the radio wave represent the rising and falling of the audible sound wave to be transmitted.[SCRO71] The advantage of FM transmission is the reduction of noise that is so common in AM transmission. The static is caused by interference or noise that competes with the original transmission at the receiver end. With AM, the *amplitude* of the carrier wave is often at lower levels (when the audible sound wave has "fallen" or entered the "lower" portion of the up and down wave). The lower amplitude is created by lowering the power of the transmission. Thus, it is easy for static to creep in when noise interferes with this portion of the audible wave (low amplitude portion of the radio wave). With FM, the *amplitude* remains *constantly high* and the sender can broadcast at full power continuously. This difference significantly reduces the problem of noise creeping into the transmission. FM receivers are obviously quite different than AM receivers. Although changes in amplitude (caused, for example, by noise) can be ignored, a method of converting changes in frequency into the corresponding (original) variations in amplitude of the audible sound waves originally generated for transmission is needed. A device known as a *discriminator* is used for this purpose.

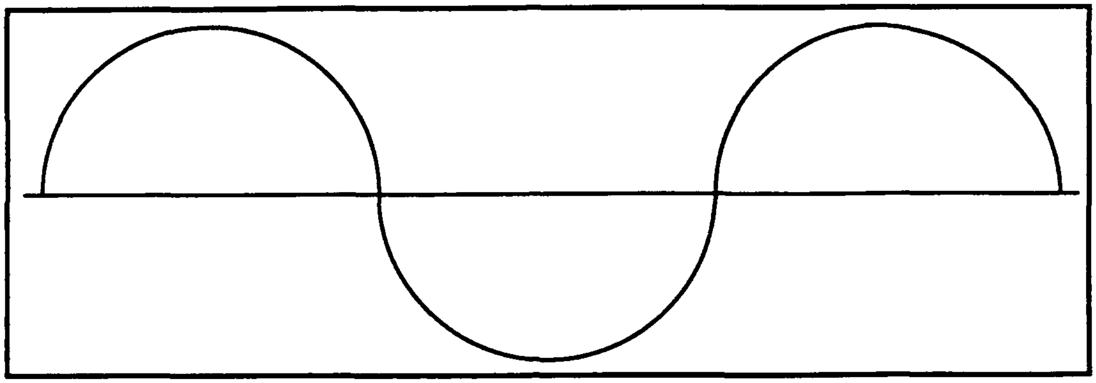


FIGURE 16. AUDIBLE LOW-FREQUENCY WAVE TO BE TRANSMITTED

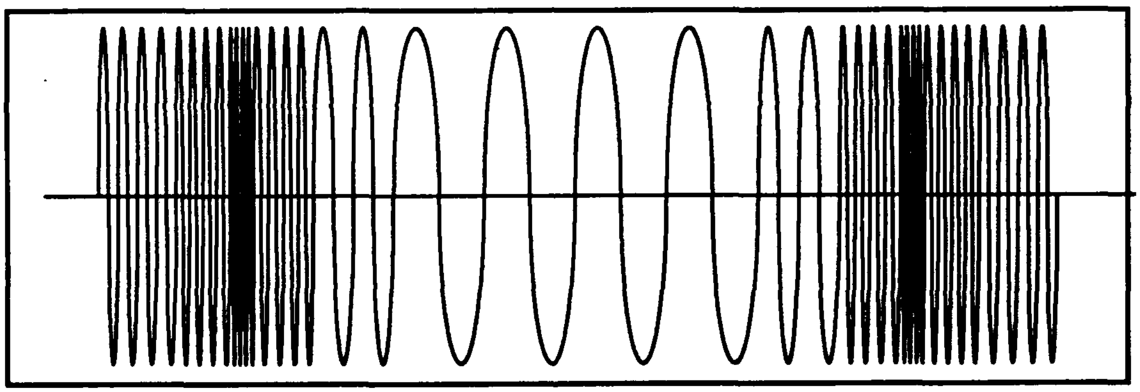


FIGURE 17. RESULTING FREQUENCY-MODULATED (FM) WAVE

IV. TELEVISION

A. EARLY HISTORY

Serious interest in what is today known as television was first evident in the 1870's. The earliest schemes for television drew on the technology previously developed for the telegraph. Both facsimile telegraphy and the synchronization schemes developed for the multiplexing of telegraphic transmissions were easily adapted by the early pioneers of television.

The primary catalyst for the early television schemes was the discovery of certain properties of the element known as *selenium.* Selenium had been commonly used as a high (electrical) resistance material in the maintenance and testing of telegraph lines. In 1873, it was discovered that the resistance of selenium was inversely proportional to its exposure to light. Thus, when high levels of light were present, selenium acted as a good conductor of electricity and, in darkness, as a good resistor (poor electrical conductor).

Soon after the discovery of selenium's photo-electric properties, several schemes were developed for transmission of images over great distances using electric wire. Some techniques involved creating large matrices of selenium cells that would be exposed to the image to be transmitted. Each cell would be connected to an electrical circuit and the amount of light to which it was exposed would determine the amount of electrical current that would flow through the circuit to a distant point where a light source would be powered. (A distant matrix of light cells were also required on the receiving end.) While this scheme involved a multiwire circuit (one for each cell), other techniques used a single wire circuit and synchronization techniques (similar to multiplexing in telegraphy) to achieve the same result.

Unfortunately, the selenium technology of 1880 proved to be a dead-end for television. The problem involved the lack of speed with which selenium cells responded. The long recovery time for selenium to regain its resistance prevented several frames of pictures from being transmitted in rapid sequence. Indeed, nearly every early scheme involving selenium was capable of producing only a single picture and, in order to transmit that image over a *single* wire, required a spiral or other type of synchronizing scanning technique. However, by 1880 research in the field of what was to become television exploded after Graham Bell made public his experiments with a telephonic system that employed voice-modulated light (the "photophone").[SHIE77] Additionally, selenium research continued which resulted in more sensitive forms of the substance being discovered.

While many schemes were proposed for using selenium to transmit images, nearly all were nothing more than telegraphic facsimile systems. By the end of 1883 Paul Nipkow, a student in Berlin, developed what is generally considered to be the first working *mechanical* television system.

B. MECHANICAL TELEVISION SYSTEMS

Paul Nipkow's scheme for what he called the "electric telescope" was a relatively simple combination of existing technology. Using a rotating perforated disk to scan images, light passing through the disk was converted into electric current in proportion to its intensity. The resulting current was placed into a (potentially) long-distance circuit and, at the receiving end, converted into light of varying intensity. This light was in turn viewed through an identically synchronized rotating disk on the receiving end. Nipkow proposed 10 rotations per second in his synchronized scanning/viewing disks (each of which had 24 perforations). Figure 18 below illustrates what the disk looked like. The resulting image thus had 24 horizontal definition lines that were refreshed every tenth of a second. This can be compared with today's 525 scanning lines each of which is refreshed 30 times a second.

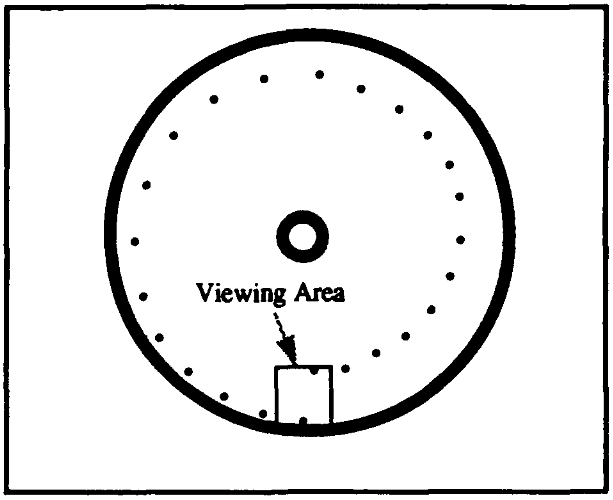


FIGURE 18. THE NIPKOW SCANNING DISK

The illusion of moving images was finally possible. Unfortunately for Nipkow, he never took advantage of his invention. The patent he received in 1885 lapsed a few years later without attempts to commercialize the device. In fact, many years elapsed until an actual working model of his device was constructed (see Figure 19 below). Modifications of Nipkow's mechanical television continued to be the state of the art until 1933 when the picture tube rapidly replaced the mechanical version of television.

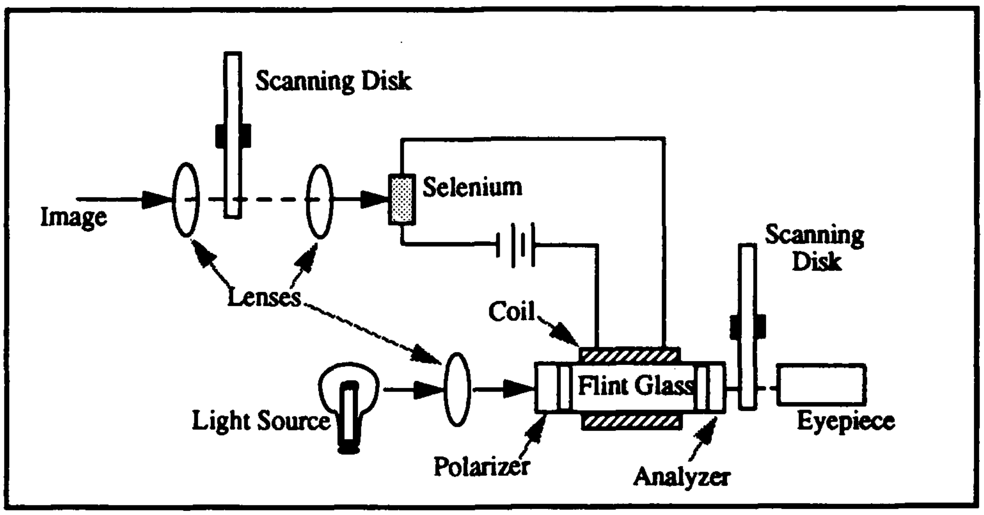


FIGURE 19. PAUL NIPKOW'S ELECTRIC TELESCOPE OF 1884[SHIE77]

It is interesting to note that among the world leaders in mechanical television systems and broadcasting were the Soviets. While such a fact might seem quite surprising in view of the fact that the Soviet Revolution had occurred during the early evolution of this technology, it was the interest of Lenin himself that made this possible. Moscow itself had regular television broadcasting to the general public (at least those with access to a receiver) by 1931. At the same time, only experimental broadcasts were taking place in Great Britain, Germany, and the United States.[SHIE77]

C. ELECTRONIC TELEVISION SYSTEMS

By the early 1930's it had become evident that mechanical television systems were rapidly approaching their physical limits. The problems associated with synchronization at ever higher speeds and dust clogging ever smaller apertures in scanning disks (as the number of scanning lines increased) were becoming obvious. The time for another, non-mechanical, approach had arrived. Application of vacuum display tube-based screens and a new type of camera made an electronic approach possible.

1. Cathode Ray Tube (CRT)

Cathode ray tubes are vacuum tubes that use the cathode (negatively charged electrode) to discharge a stream of electrons onto a fluorescent screen. The impacting of the electrons upon the fluorescent screen creates visible light--making the point of impact visible. Further, by placing an external magnetic field in the vicinity of the beam of electrons, the electrons can be directed to specific areas of the screen. By adjusting the intensity of the magnetic field, the stream of electrons can be made to scan the entire screen in a precise, synchronized manner. The ability to bend (or deflect) these electron beams was a significant technical challenge. Early television receivers using CRT technology had relatively tiny screens yet the tubes were very long (measured from cathode to screen). This was because the angle that the electron stream could be accurately deflected was relatively small in the early years. As the ability to accurately bend these streams increased, the tubes became flatter. Even today, a large, flat, and thin picture tube eludes developers.

The earliest type of cathode ray tube was developed in 1897 (the Braun tube). While quite limited in expected life and ability to focus, it became widely used for scientific purposes such as laboratory measurements. CRT technology continued to improve over the next three decades to the point where its power consumption, ability to focus, and life-span were adequate for incorporation into television reception schemes.

Today's television tube technology has been easily adopted by the computer industry in the development of display screens and monitors. Many modern medical and scientific instruments also employ modern CRT technology that was first developed for use in television.

2. Electron Cameras

The earliest electron cameras were developed in the 1920's specifically for use with television. The purpose of such cameras was to convert images into a series of electrical variations that can then be transmitted to receiving stations for conversion into their original form. While the technology is relatively complex, the principal parts of such a camera can be described with the help of the Figure 20 below.[SHIE77]

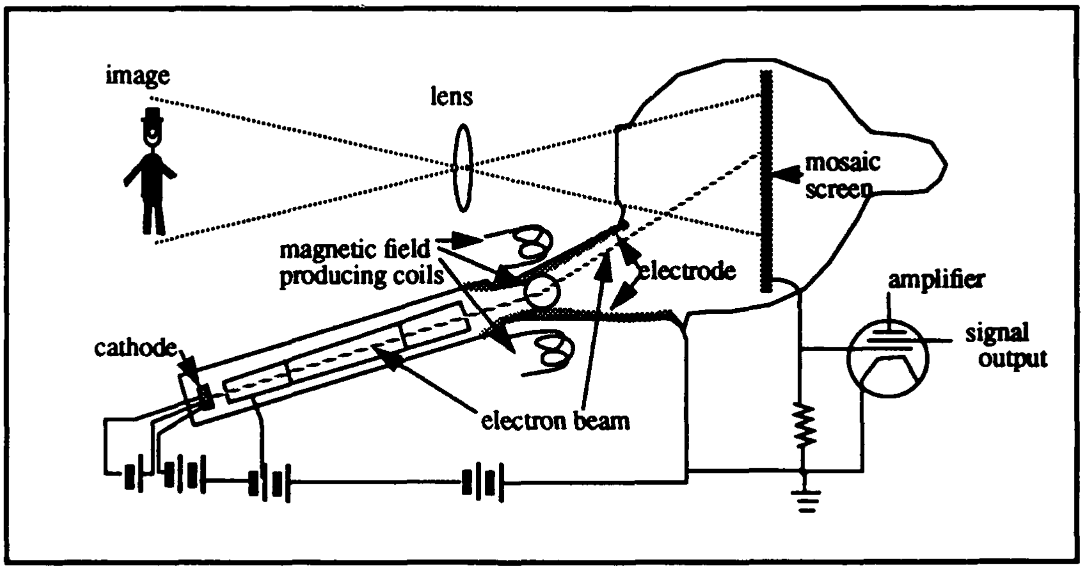


FIGURE 20. THE EMITRON TELEVISION CAMERA

The camera itself consists primarily of a large vacuum tube. Inside the tube is a mosaic screen upon which the image to be transmitted is projected (with the help of an external lens). An electron beam is generated by the cathode depicted in the bottom left portion of the diagram. Another, positively charged, electrode consists of a silver coating on the inside of the vacuum tube (connected to its power source via a wire that pierces the tube). Near the positively-charged electrode are several wire coils that produce magnetic fields capable of deflecting the electron beam onto desired areas of the mosaic screen. Specifically, they are caused to scan the surface of the screen in a series of parallel horizontal lines.

The mosaic screen is composed of mica and silver that allows electrons to be released in proportion to the amount of light falling upon it from the image projection. When the electron beam is in contact with that point on the screen, the deficiency of electrons is replenished by the beam and, in turn, fewer electrons from the beam make their way out of the tube via a wire connection that feeds into an amplifier (or series of amplifiers) prior to being useful for transmission. All the information necessary to recreate the image just scanned is present in the electrical flow produced.

3. Pioneers of Electronic Television

Just as radio had its Marconi and other significant pioneers, *electronic* television has two men that are often associated with its earliest work.[SHIE77]

a. *Farnsworth*

Philo T. Farnsworth began using a modified cathode ray tube for both the transmission and reception of images prior to 1929, the year he first demonstrated the technology. He joined the Philco Corporation in 1931 where he worked for the next couple of years developing the technology further, to include an experimental broadcasting station in Philadelphia.

b. Zworykin

Dr. V. K. Zworykin, another American, led early electronic television research at Westinghouse Research Laboratories. He demonstrated an all electronic television in 1924. He joined RCA in 1930 where he continued to advance the technology.

D. COLOR TELEVISION

Schemes for the incorporation of color into television were first demonstrated in the late 1920's with mechanical television systems. Most involved the addition of color filters in the system's spinning disks, that also required a high degree of coordination and synchronization. Several more modern schemes for color television were developed by the early 1950's and, by 1953, the Federal Communications Commission (FCC) had promulgated standards for color television. Perhaps the most significant aspect of the FCC decision was to keep color broadcasts compatible with the existing black and white systems to prevent the obsolescence of existing receivers.

Color transmission of images usually involved splitting the image into its three primary colors (red, green, and blue) with the aid of color filters. The resultant primary colors were then captured on camera, or, in some schemes, with three cameras. The signal information was then sent as before to a receiving television. For color reception, modifications to existing television tubes were necessary. Specifically, the phosphor elements of the screen (also known as picture elements or *pixels*) were modified to emit one of the three primary colors and the three types of phosphor elements were equally distributed to specific areas of the screen. The electron beam generated by the tube's cathode is precisely aimed at the specific area and correct phosphor color element to remain in synchronization with the broadcast signal. An early approach to color tubes involved alternating red, green, and blue phosphor element lines in on the display area of the screen. This scheme required that each image be scanned three times (a *triple interlace*) once for each color.

Today's common receiver tubes are known as *shadow mask tubes*. The tubes usually have one electron gun (cathode) for each color--three in all. Such tubes arrange each of the three colors of phosphor elements in groups of three--making equilateral triangles. The triangles are then placed upon each scan line in long, horizontal rows upon the tube's screen. The orientation of each of the color phosphors is such that they can be precisely hit by the appropriate electron gun.

E. HIGH DEFINITION TELEVISION (HDTV)

As the name implies, HDTV improves upon the existing television resolution to further improve the quality of the images transmitted and received. Unlike the first three generations of television (mechanical, electronic, and color), the technology to make HDTV possible arrived *before* the demand for the technology did. This is perhaps due to the fact that an increase in resolution beyond the current standard is almost unnoticeable on display screens smaller than those with at least a 25-inch diameter. However, with the advent of "big-screen" TV, demand for higher quality images is considered by many in industry to be inevitable--given enough time.

Other types of enhanced receiver technology also exist [BELT91]. *Improved Definition Television* (IDTV) receivers have been in production for three years and the signals they receive and display are identical in every respect to the existing NTSC standard color signals. The only differences in IDTV lie with the receiver. Specifically, the analog signals are stored in a small amount of receiver memory after being digitized and enhanced by the processor present in the receiver. The resulting pictures are displayed approximately 60 times per second.

Enhanced Definition Television (EDTV) receivers should be in mass production within the next four years. Like IDTV, EDTV would utilize internal receiver-based processors to enhance, digitize, and store the signals received prior to display. Unlike IDTV, EDTV would not necessarily be restricted to the 525 scan lines used in the NTSC standard (the bandwidth for each channel, however, would remain unchanged at 6 MHz).

High Definition Television will use a different broadcasting format along with advanced processors to deliver twice as much information to the receivers as before. The increased display information will require more memory and processor capability than IDTV or EDTV.

1. Early HDTV Developments

The United States military and other organizations have been funding and performing research in the area of advanced display monitors for years. HDTV in the past had come to mean simply the *next generation* of television. No specific standards or specifications had emerged. Realizing the economic significance of HDTV, the U. S. Government has become highly active in the process of developing standards for HDTV in this country. There are far-reaching implications for the standards that will be adopted. First of all, the patented technology that is eventually adapted will, over the life of HDTV use, generate hundreds of millions of dollars in royalties. Obviously, it is in the interest of any country's trade balance to adapt domestically developed standards. Secondly, the processor and memory (micro*chip*) content of each HDTV receiver is expected to exceed 25% of the value added in manufacture. With such enormous potential sales, it is also in the interest of the government to steer as much of the HDTV business to domestic chip producers (this is one reason the U.S. department of defense funds such chip research by domestic chip producers).

2. Current Trends

Timing of standards adaptation is critical for many reasons. HDTV standards adopted too early will stifle further research into the area as broadcasters, receiver manufacturers, and others begin full scale transition to HDTV component production. Standards adopted too late risk allowing industry to push their own de facto standards that, if well liked by the public, could be very hard to change by official decree. In this country, the government (specifically, the Defense Advanced Research Projects Agency or DARPA) has opened up a competition among contracting teams to design a HDTV technology standard. The FCC is scheduled to decide on the standards for HDTV sometime in 1993. For now, the FCC has only decided that HDTV will be compatible with existing television receivers (decided in 1988). Thus, HDTV broadcasts will be viewable by those with non-HDTV sets. Another FCC decision calls for the adoption of a *digital* standard (which, because of decision number one, can be processed by current analog-based receivers).

Other countries have had problems with rushing their HDTV to market. Japan, for example, has adopted an analog-based HDTV system known as Hi-Vision in 1984. The cost of receivers exceeds $30,000 each and programming in this mode has been limited. There are also shortfalls in the video-compression technique, known as MUSE, that allows the extra video and audio information to fit within its allowed channel bandwidth. Like the European systems, Japanese HDTV also relies on direct-broadcast satellites to transmit the HDTV signals.[BELT91]

HDTV technology, however it evolves, will quickly spill over into other industries and technologies. Its display technology will adapt relatively easily to computer monitors (as well as monitors for aircraft, vehicles, and consumer electronics). The signal processing and compression technology refined for HDTV microchips will no doubt be of use to the video telephone industry. Numerous other cross-applications exist.

F. REGULATIONS AND STANDARDS FOR TELEVISION

Like radio, the need for standardization and regulation of the television industry was quickly recognized. Early work on standardization was carried out by the Radio Manufacturers Association (R.M.A.). The R.M.A. had been highly successful in the establishment of radio standards in the 1920's. By 1929, the R.M.A. Committee on Television was active in their premature attempts to standardize the industry. It was not until 1936 that the R.M.A. proposed a number of standards as well as bandwidth allocation recommendations to the Federal Communications Commission (FCC). Such standards included, among other things, the channel width, scan rates, lines of resolution, and aspect ratio of the picture to be broadcast. These proposals formed a solid framework for commercial standards that were eventually developed five years later. On April 30, 1939, the National Broadcasting Company inaugurated its limited broadcasting of television (along with sales of receivers to the public) at the opening of the New York World's Fair[SHIE77].

In 1940 the National Television System Committee (NTSC) had been established to reevaluate the rapidly changing technology and proposed standards for an industry that was about to become a commercial reality. The committee was comprised of representatives from the radio and television broadcasting and manufacturing industry. Its president was appointed by the R.M.A. and the recommendations of the NTSC were submitted to the FCC by the board of directors of the R.M.A. Such self regulation of the radio industry had worked incredibly well two decades earlier. At a national level, the same proved to be the case for the television industry. By 1941, comprehensive standards, proposed by the NTSC, were adopted by the FCC. This cleared the way for the full scale commercialization of the industry. While World War II put a hold on television commercialization, the standards allowed for the mass production of standardized receivers and licensing of a growing number of broadcasters in the years following the war. In 1953, standards were adopted for color television as well. Unfortunately, no international standards for television were ever adopted. As a result, broadcasts between many countries cannot succeed without special adaptors or other non-standard receiving equipment.

1. United States (NTSC) Standards for Television

Guided largely by the NTSC, the FCC has adopted a series of standards for television within the U.S. A few of the more significant aspects of these standards are summarized below.

a. Resolution

The quality of the images received is directly related to the resolution of the image transmitted. Resolution is increased as the number of scan lines increases. Evolving de facto standards prior to 1941 called for 343 lines in 1935 and increased in 1947 to 441 scan lines. The standards adopted in 1941 called for the current 525 scan lines used by today's (non-HDTV) receivers. (It should also be noted that in 1950, a 405 line standard for *color* television was briefly adopted--it was rescinded three years later and replace with the 525-line standard.) High Definition Television (HDTV) Standards are still evolving. Since the early 1930's, the refresh rate of the images has been 30 times per second.

Of today's 525 scan lines, only 483 are actually visible on the screen. The time used to reposition (retrace) back to the top of the screen, after scanning from top to bottom, makes it impossible to use the remaining 42 lines.

To avoid flicker, standards also call for the interlacing of images. By interlacing the scan lines, the odd scan lines are scanned from top to bottom of screen, then the even numbered are scanned from top to bottom. Interlacing requires that the scanning be performed 60 vertical cycles per second. Since two vertical scan cycles are necessary for a complete image, the image refresh rate is 30 times per second. One can compare this rate with that of modern movie films that use 24 frames per second.

b. Modulation

As discussed in sections III.D. and III.E. of this paper, there are two modes of modulating the signals used in radio--amplitude modulation (AM) and frequency modulation (FM). Television broadcasts use both. While FM is used for the sound portion of television broadcasts, AM is used for transmitting the image information. FM was briefly tested for use in image transmission, but it was discovered that FM images were more easily disturbed by *multipath* interference. (Multipath interference is caused by signals being reflected by natural or man-made objects in such a way as to create *multiple paths* between the broadcasting and receiving antennas.)

The standards adopted in the United States call for *negative polarity modulation*. In other words, the more light present in an image, the less carrier amplitude is used in the analog transmission of this fact.

c. *Channel Bandwidth*

The increase in resolution to 525 scan lines in 1941 made it necessary to increase the bandwidth of each channel. The video information requires slightly more than 4 MHz of bandwidth and the audio portion requires and additional, relatively small, fraction of this. In total, 6 MHz of bandwidth has been adopted as the standard. This allows for the additional information about color as well as adequate space for adjacent channel separation.

2. International Television System Standards

While most of the western hemisphere and the far east have adopted the television standards of the United States, significantly different standards exist for the rest of the world. Indeed, the only significant standard adopted world-wide is the 2-to-1 interlace used to prevent flicker. A primary reason for many of the differences relates to the power supplies found elsewhere. Most of the rest of the world uses lower frequency power. Specifically 50 Hz power is more common than the 60 Hz power found in the U.S. As a result, the refresh rate is only 5/6th of that found in the U.S. (25 times per second vs. 30). The table below summarizes the more significant standards.[SHIE77]

TABLE 3: WORLD TELEVISION STANDARDS

Standard	British	Western Hemisphere & Far East	Western Europe	Russian & Eastern European	French
Scan Lines	405	525	625	625	819
Channel Width	5 MHz	6 MHz	7 MHz	8 MHz	14 MHz
Sound Modulation	AM	FM	FM	FM	AM
Picture Frequency	25 / sec.	30 / sec.	25 / sec.	25 / sec.	25 / sec.
Polarity Modulation	positive	negative	negative	negative	positive

Conversion between such standards used to involve making second generation copies of television pictures that used another format. Since the 1950's however, electronic converters have been developed that use no optical interface for the conversion. The result is a clear, first-generation-like copy of the original broadcast.

V. SPECIFICATION AND ANALYSIS OF A CELLULAR RADIO NETWORK

Cellular radio is a method of providing wireless telephone service to (primarily) mobile users of the public telephone system. Cellular radio divides up metropolitan areas into *cells*, each of which contains an antenna capable of transmitting and receiving low-power radio signals within its assigned area--primarily the cell. The antennas are connected via land line to a mobile telephone central office (MTCO) which coordinates the switching of circuits as necessary. The MTCO is also responsible for tracking the location of the mobile telephones in order to continuously use the antenna located within the same boundary as the mobile telephone in use. Using a portion of the UHV band and relatively low power levels, cellular telephone makes efficient use of limited bandwidth via space division multiplexing (SDM).

A. OPERATIONAL OVERVIEW

The operation of a typical cellular radio network can be illustrated with the following example and diagrams. As mentioned above, a typical metropolitan area serviced by cellular radio is divided into cells. A single antenna located within each cell maintains radio contact with the mobile subscribers in its cell. Figure 21 shows how a metropolitan area may be broken up into cells. The cells themselves are depicted as hexagons and serve as the approximate boundaries of the service area for each antenna.

A variety of UHF frequencies are available for use by the cellular network. To avoid interference, *adjacent* cells avoid using the same frequencies. This, in combination with low power broadcasts, allows efficient use of available bandwidth since cells not adjacent to each other may reuse similar frequencies. One exception to this rule involves call initiation. As calls are initiated by cellular phones, a single off-hook frequency can be used by the callers to signal their intention to make a call. This frequency is monitored by all antennas within the cellular network. As the off-hook signal is detected by multiple antennas, the power level of the signal detected is reported to the MTCO. The MTCO compares the power levels reported and determines which antenna will take over the call. This will generally be the antenna located within the same cell as the cellular phone. The supporting antenna then signals to the cellular phone that a dial tone can be obtained on a particular frequency. After the cellular phone changes to the assigned frequency, the call proceeds as a normal telephone call would. If, during the course of the call, the signal becomes weak enough to indicate the cellular phone has left the antenna's cell, the MTCO can intervene to hand off the call to an adjacent antenna which is now closer to the mobile telephone. Such hand-off will again require power-level measurements and frequency switching similar to that which takes place during the off-hook portion of call initiation.

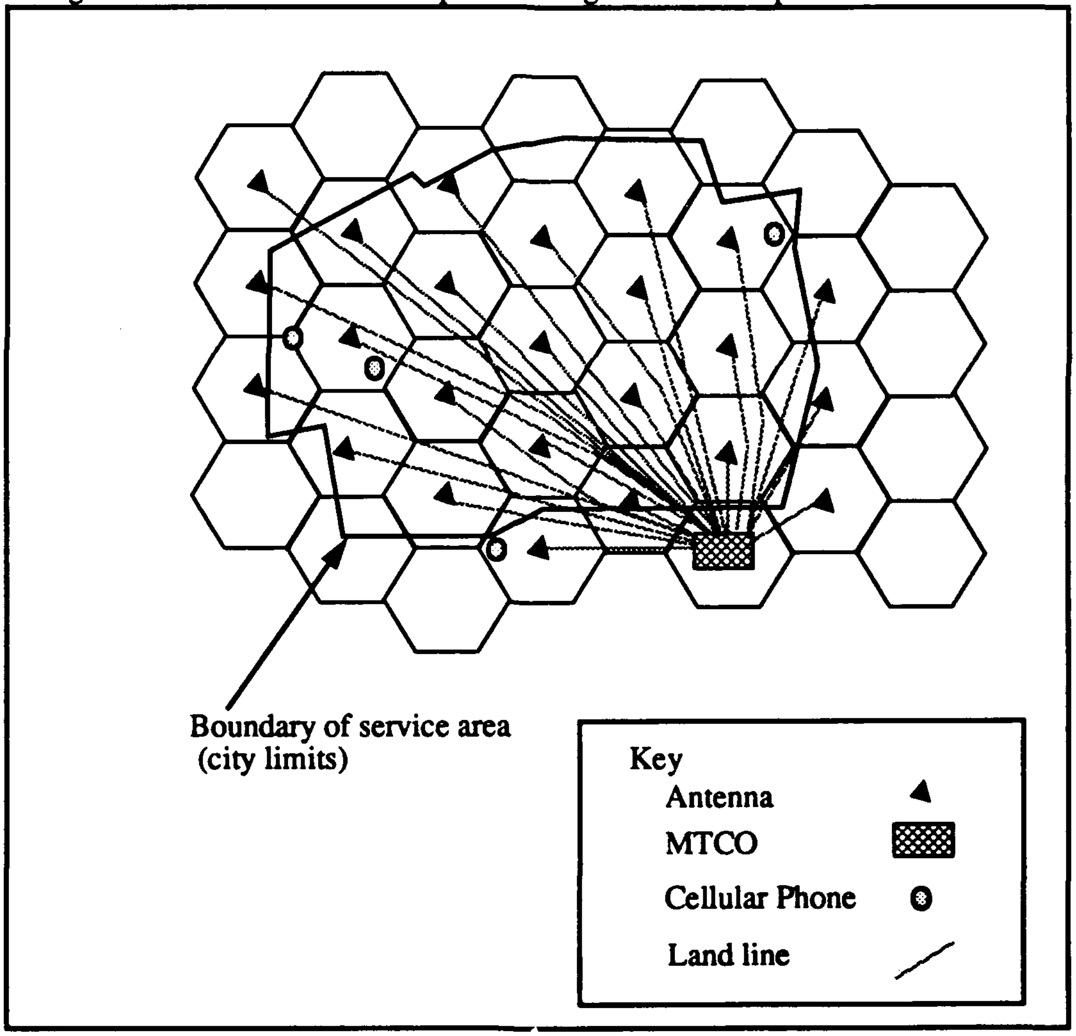


FIGURE 21. CELLULAR NETWORK

B. CELLULAR RADIO PROTOCOL

Although the protocol used by cellular radio companies can vary, one possible protocol is outlined below. First a brief description of the system components is given. Then a formal description of the protocol using a formal model is presented. Finally, an analysis of the network is done, which proves certain critical correctness properties.

1. Cellular Radio Components

For purposes of this analysis, only the mobile telephones, the antennas, and MTCOs will be considered. Some cellular systems, especially rural ones, may involve satellites and nearly all systems interface with the public telephone network. However, the protocol analysis considers only the cellular radio portion of the network.

a. Cellular Telephones

The cellular telephones are the mobile devices which subscribers to the cellular service use. While frequently found in automobiles, they may be hand-held (pocket telephones), or, near the coastline, marine based. In some lesser developed countries, such as India or Mexico, cellular phones even replace regular telephones in residences and businesses. (This is because of the long waiting times involved in establishing normal telephone service.)

Cellular telephones are capable of low-power radio transmissions which can be accurately received by the servicing antenna of the cell in which they are located. For two-way communication, they are also able to receive and adequately amplify the broadcasts from the servicing cell's antenna. The most sophisticated aspects of such telephones are the integrated circuitry and microprocessor capabilities. These allow the phones to quickly change frequencies when directed by the servicing antenna. They also automate the encoding and broadcast of the identifying subscriber account number for billing and incoming call purposes.

b. Antennas

The antennas which support the cellular network are located in the vicinity of the center of the cell which they service. They too are capable of low power radio transmission and reception to and from cellular telephones located (primarily) within their cell. They are capable of transmitting and receiving on a large number of UHF frequencies simultaneously. The antennas are equipped with special measuring equipment which allows them to measure the power of each incoming transmission. This information is reported to the MTCO. Their land-line connection to the MTCO allows them to serve as extensions of the central office. When the antennas become overloaded with call volume, or when telephones being serviced move out of their range, the MTCO will direct the hand off of calls to adjacent antennas.

c. Mobile Telephone Central Office

The MTCO serves two functions. First of all, it manages the network of cells and the activity of the antennas to which it connected. Secondly, it serves as the interface, or gateway, to the public telephone system for calls requiring any portion of their circuit outside of the cellular network.

In managing the cellular network, the MTCO performs a variety of functions. When calls are originated by mobile telephones, the MTCO determines which of the antennas is closest to the telephone and assigns it the responsibility of servicing the call. This determination is made based upon the antenna's power level measurements of the incoming radio signal. During cellular telephone calls, the MTCO is responsible for coordinating the hand-off (or transfer) of calls which, due to the changing location of the telephone or high volume of traffic being serviced by an antenna, need to be transferred to an adjacent cell's antenna. Most importantly, the MTCO is responsible for real-time frequency management.

2. A Protocol for Cellular Radio

The FSMs which follow describe the various states the three components of the cellular network can be in.

a. Cellular Telephone FSM

The cellular telephone can be in one of nine states. State zero represents its dormant state--the phone is not in use for any aspect of an incoming or outgoing call. When an outgoing call is initiated and the handset is removed, an off-hook signal is generated which will be picked up by one or more antennas (if within range). As this off-hook signal is generated, the FSM moves into state one.

From state 1, two things can happen: first, the caller can hang up prior to the phone switching to the assigned frequency or, second, the assigned frequency can be received by the mobile phone. If the caller hangs up from this or any other state, the FSM moves back to state zero. If the new frequency is received by the cellular phone, the phone will move to the assigned frequency--this is represented in the FSM by state two.

While in state two, the telephone should momentarily receive a dial tone. This is represented in the FSM by moving to state three. The dialing of the desired phone number moves the FSM's state to state four.

From state 4, three possible transitions exist. As from all states, the caller can hang up and return to state zero. However, as the network attempts to establish a circuit for the call, it may or may not be successful. If unsuccessful, the FSM moves to state six (busy signal). A successful circuit establishment will result in the ringing of the distant telephone and subsequent conversation or data exchange. This is state five of the FSM.

The FSM remains in state five until the caller hangs up or moves out of the servicing antenna's cell. In the case of the caller moving out of the cell during a conversation, the MTCO will direct a hand-off of the call to an adjacent cell and the cellular telephone will be notified of a new frequency to change to. This new frequency notification during a call is represented in the FSM as state seven. The process of switching frequencies is shown in the FSM as moving back to state five.

Incoming telephone calls involve only states 0, 5, 7, and 8. While in state zero, the cellular phone monitors a universal incoming call frequency listening for its identification number and frequency to switch to. In the event of an incoming call, the cellular phone automatically switches to the assigned frequency and begins to ring (state 8 of the FSM). If the call is subsequently picked up, the FSM moves to state five. Otherwise, when the ringing eventually stops, the FSM returns to state zero.

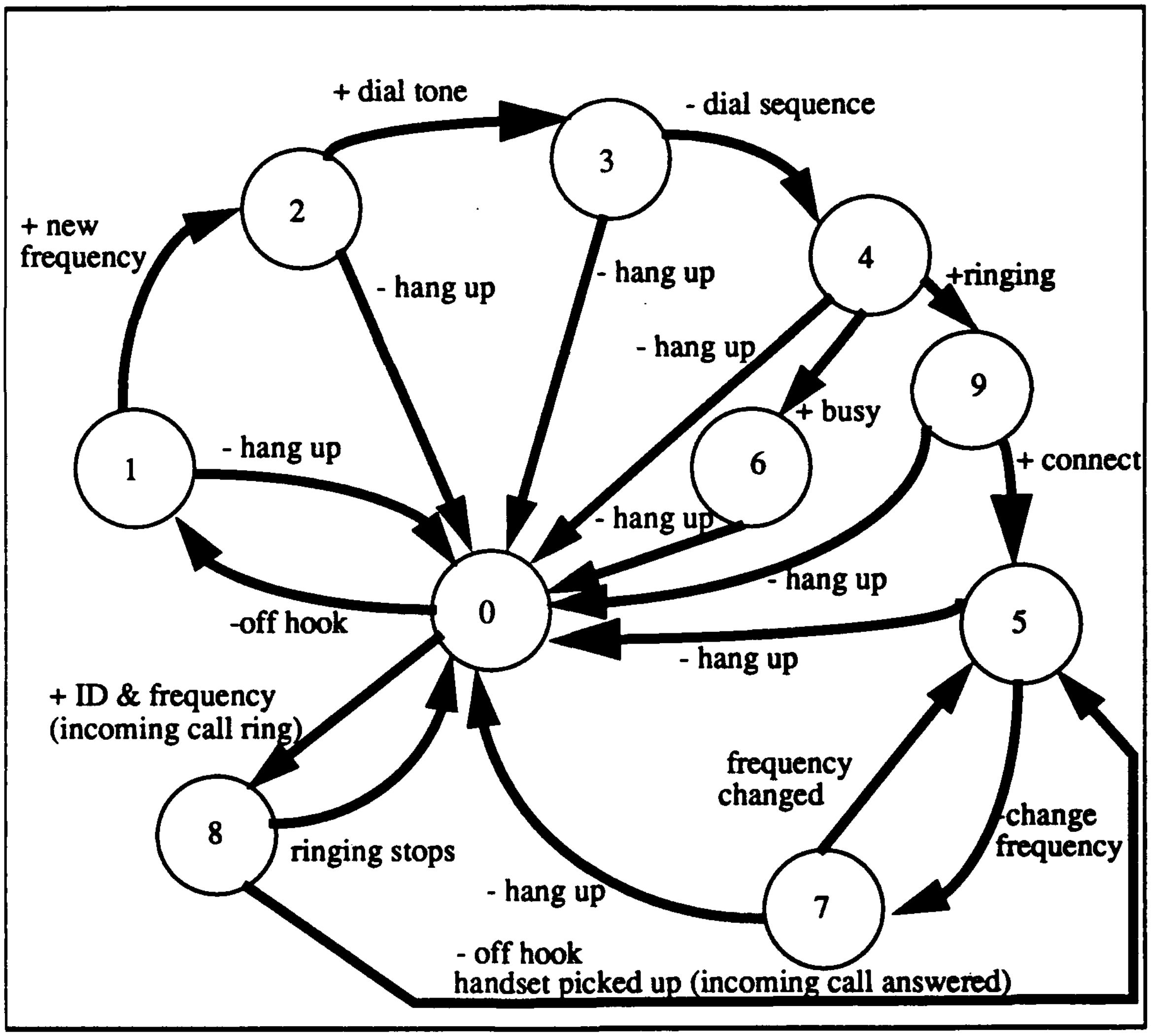


FIGURE 22. CELLULAR TELEPHONE FSM

b. Antenna FSM

On dramatic difference in the FSM for the antenna is that *a separate machine exists for each channel the antenna can operate on.* If 100 pairs of frequencies are being used by an antenna, 100 FSM's exist simultaneously which represent the state of the antenna. The operation of the FSM closely parallels that of the cellular phone it is serving. Essentially the antenna acts only as a conduit for the actions directed by the MTCO.

Figure 23 is the FSM for the antenna and references to the subscriber and MTCO are abbreviated with the symbols (S) and (C) respectively. The FSM shows that the completed circuit is reached in state 12. The states 0 through 12 are associated with normal outgoing calls originating with any subscriber within the cell. An incoming call is shown by the machine transitioning from state 0 to 13, 14, and finally 12. States 15, 16, and 17 relate to the mobile subscriber leaving the cell's service boundaries and the call in progress being handed off to an adjacent cell. A similar transition is shown for incoming hand-offs by the arrow connecting state 0 to state 12 directly. Although not shown, it should be made clear that a transition to state 0 is possible from any state via a "+ hang up" transition.

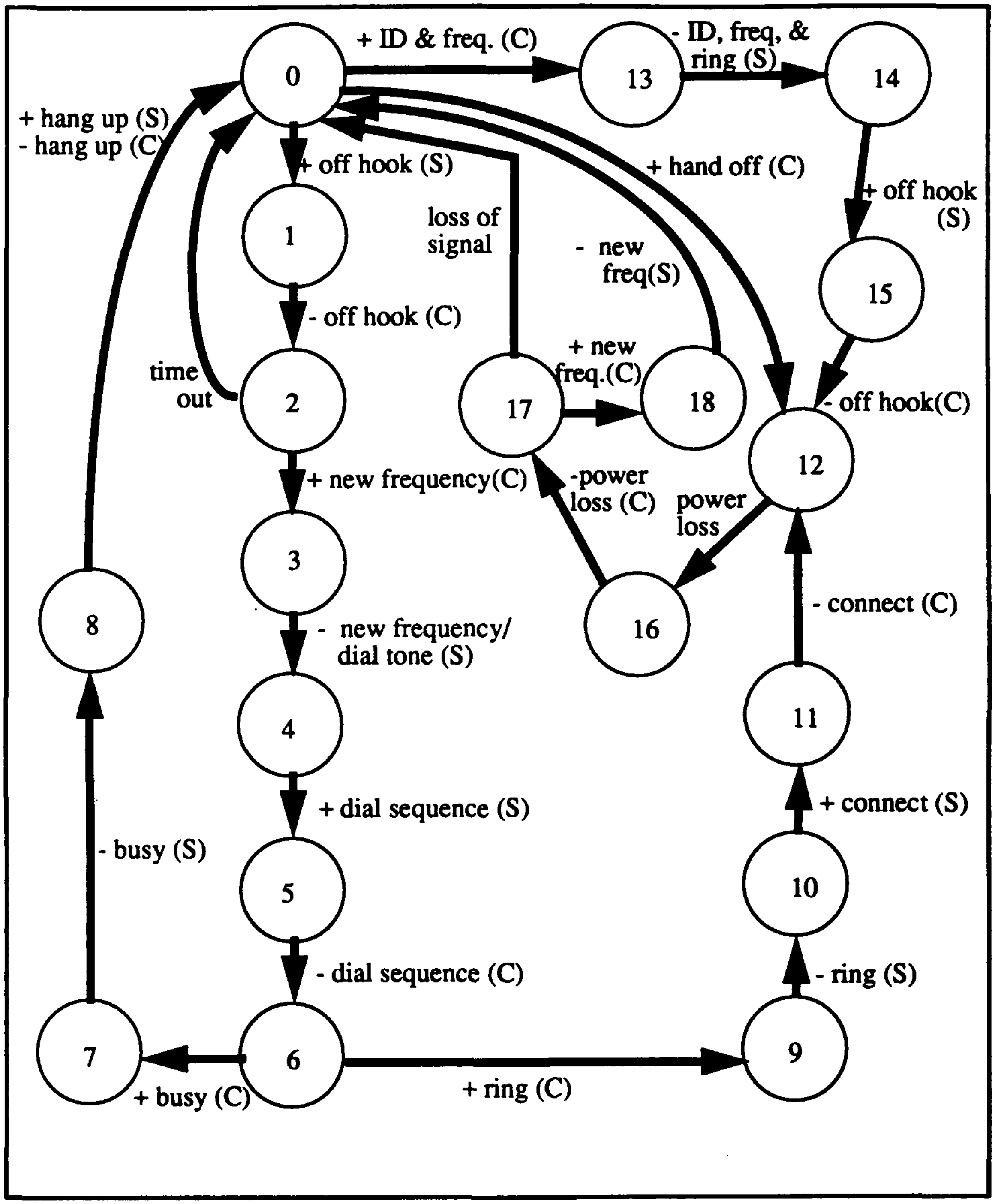


FIGURE 23. ANTENNA FSM

c. MTCO FSM

The FSM for the Mobile Telephone Central Office is the most complex. Like the antenna, a separate FSM exists for each call the MTCO is capable of handling. For simplicity, higher level functions such as frequency management or gateway activity to the public telephone system and switches is ignored. Although not shown, a "hang up" transition back to state 0 is possible from all other states.

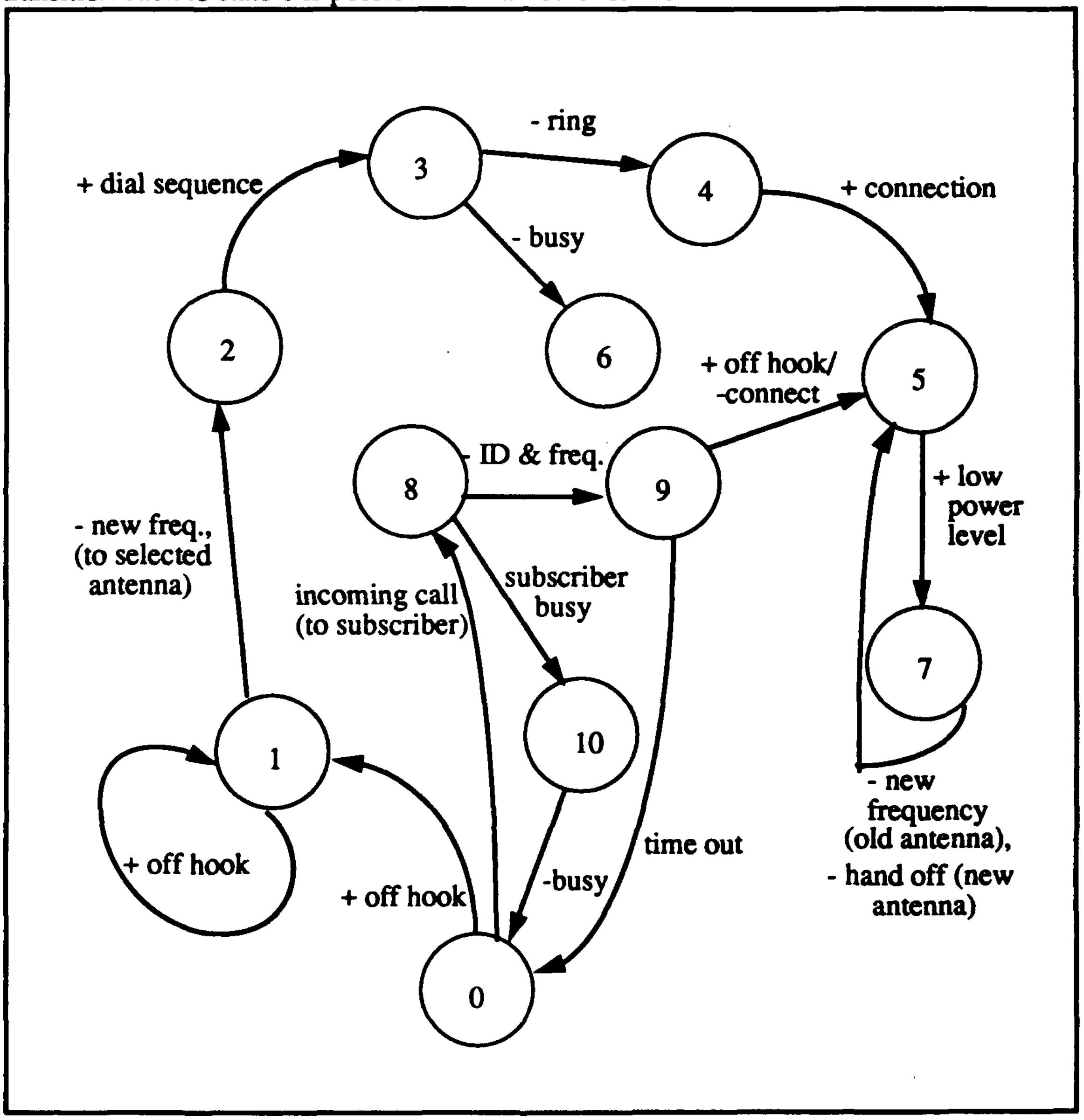


FIGURE 24. MTCO FSM

3. Analysis

The analysis of the cellular radio network protocol just described is presented via a presentation of reachability of all states without deadlock for a number of cases. The first case shown is the origination of a call from a subscriber to a destination outside the network. Secondly, an incoming call (to a subscriber) is shown. The notation used in the reachability analysis will describe the state of the subscriber (telephone), the servicing cellular antenna, and the MTCO respectively. Thus the notation $< 1, 2, 3 >$ would denote the cellular telephone FSM being in state 1, the antenna FSM being in state 2, and the MTCO FSM being in state 3. To enhance clarity, some transitions are preceded by the symbol of the machine making the transition (in brackets) and followed by the symbol of the machine impacted by the transition (in parenthesis). Thus, a description such as "[A] - off hook (C)" would indicate that the antenna, [A], is sending an off hook signal to the central office, (C).

a. Subscriber-originated Call

Figures 25 and 26 illustrate the series of states which the FSMs transition into during the course of a mobile telephone subscriber-initiated call. Initially, all FSMs are in their own state zero. When the subscriber picks up the handset to dial, an off hook signal (along with the subscriber identification number) is broadcast to all antennas within range. The antennas able to detect the signal relay this information to the MTCO along with their measurements of the subscriber's signal strength. The MTCO determines the closest antenna to the subscriber and sends it the new frequency to direct the subscriber to change to for placing the call. When the selected antenna broadcasts this frequency to the subscriber, it simultaneously broadcasts a dial tone on the new frequency. At this point in the analysis, the state of the FSMs is described by the notation $< 1, 4, 2 >$.

The subscriber's telephone, upon receipt of the newly assigned frequency, changes to that frequency, receives the dial tone, and is free to begin the dialing sequence. The sequence is picked up by the servicing antenna which relays it to the MTCO. Two possible results occur at this point--the line is busy or it will begin ringing. The MTCO transmits the result to the antenna for relay to the subscriber. The busy signal path is depicted along the lower left portion of Figure 25 and ultimately results in the subscriber hanging up and all FSMs returning to state < 0, 0, 0 >.

The alternative possibility is that the call will successfully be put through. This path is shown as the lower right portion of Figure 25. After the central office connects the call, the ring is generated and, if the distant party picks up, the connection is completed. The result of a completed connection is the arrival at state < 5, 12, 5 > by the FSMs.

Figure 26 shows the two paths a completed call can follow once established. The left branch illustrates a normal call termination via the subscriber hanging up. The right branch displays what takes place as the subscriber moves out of the cell served by the antenna initially involved in the call. As the servicing antenna notices the loss of power in the subscriber's transmission, the MTCO is notified. The MTCO determines the best alternative antenna to hand the call off to and signals the old antenna to relay the new frequency to the mobile subscriber. After doing so, the old antenna returns to state zero and a new antenna takes over the call at the new frequency--returning the FSMs to the state reached after the connection was initially established, < 5, 12, 5 >.

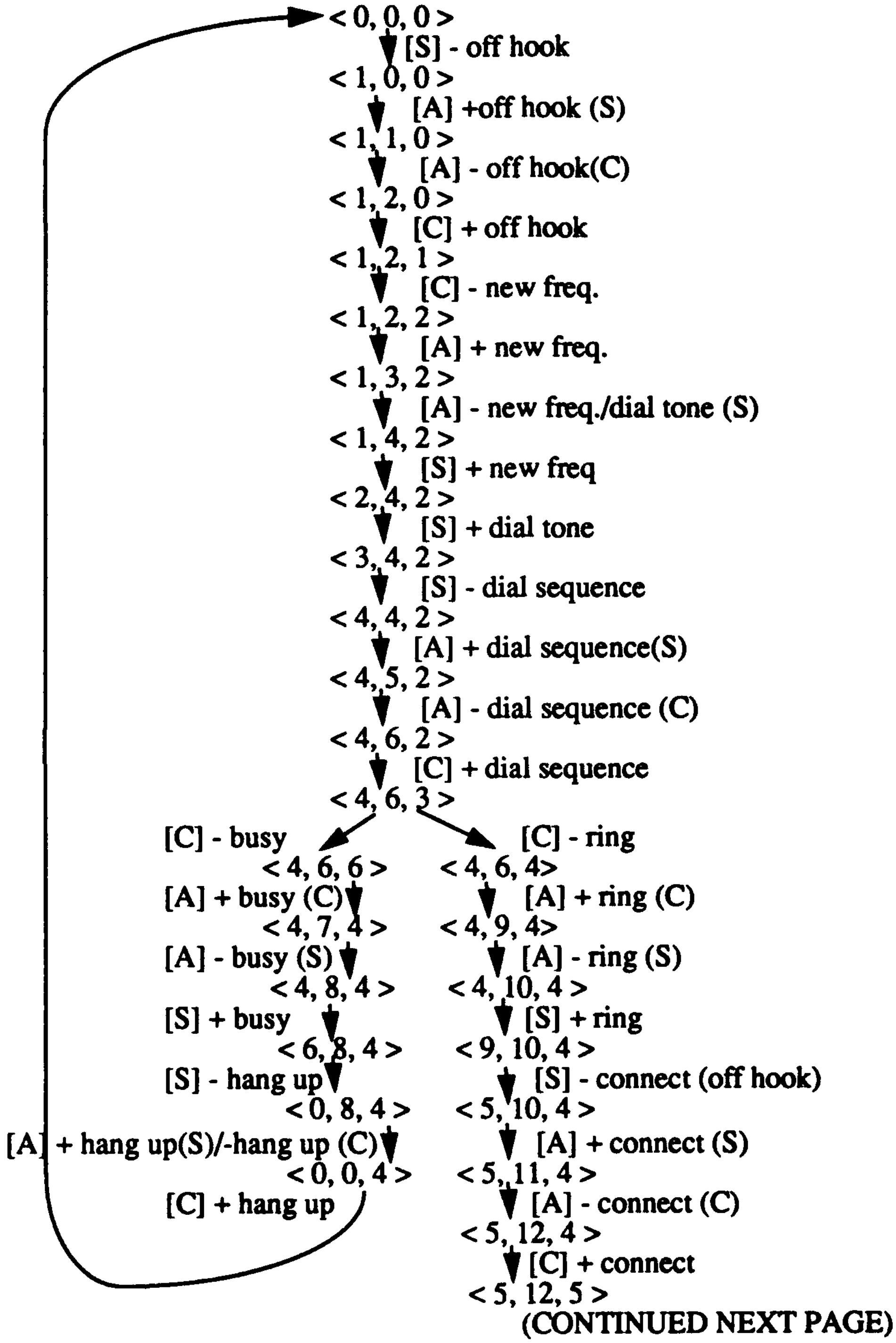


FIGURE 25. SUBSCRIBER-ORIGINATED CALL ANALYSIS

(CONTINUED FROM PREVIOUS PAGE)

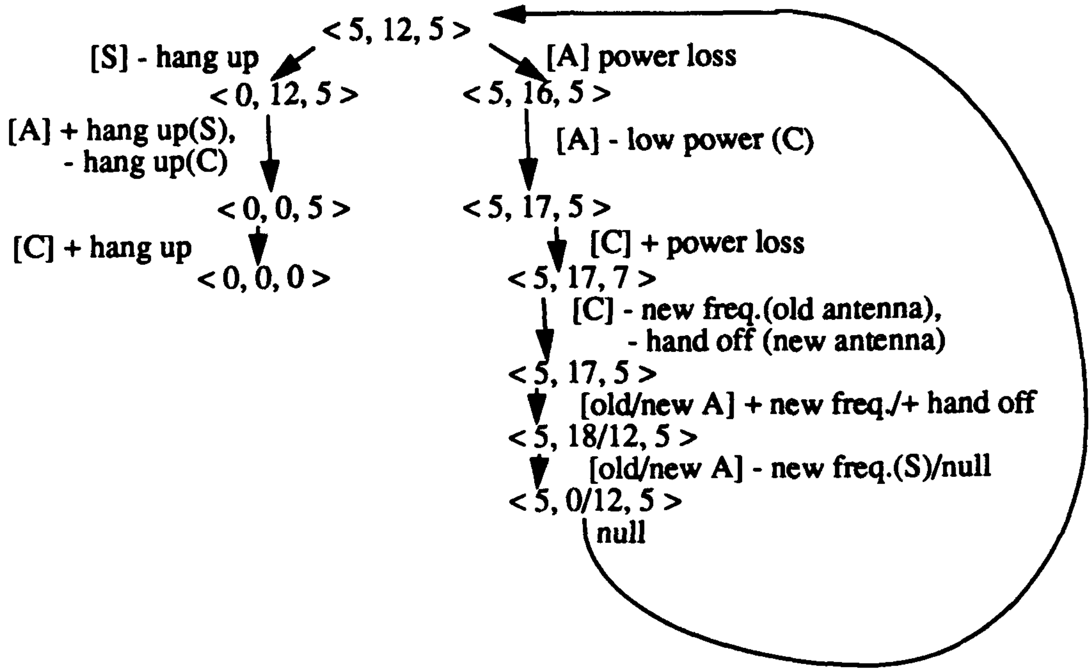


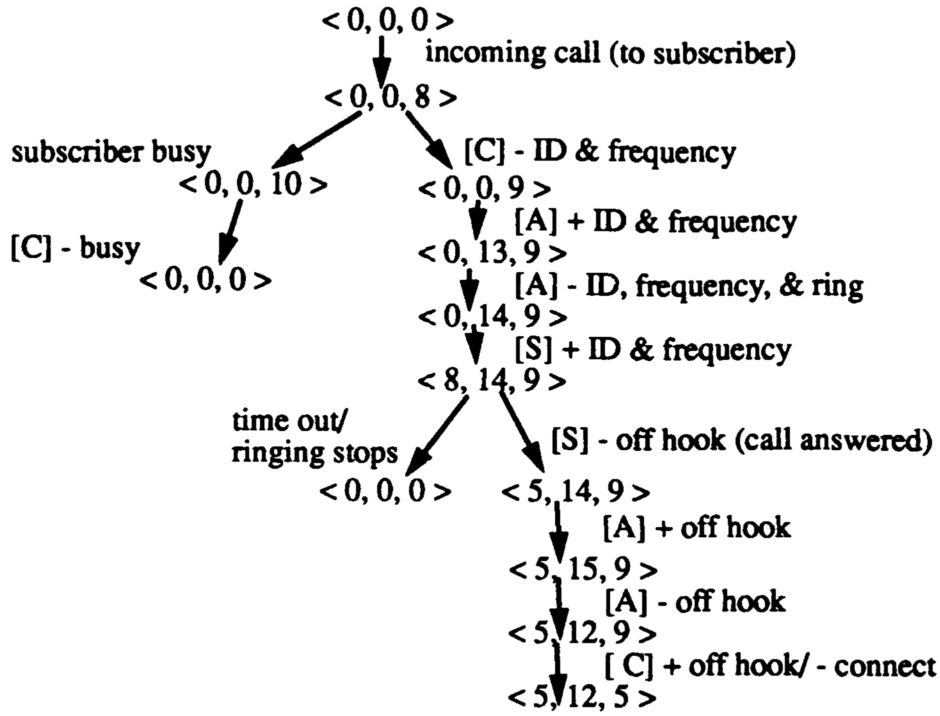
FIGURE 26. CALL HAND-OFF/TERMINATION

b. Subscriber-destination Call

Figures 26 and 27 illustrate an incoming call to a subscriber in the cellular network. The incoming call is initially shown in figure 27 when the FSMs transition from state < 0, 0, 0 > to < 0, 0, 8 >. At that point the MTCO determines from available information if the subscriber being called is currently using their telephone. If the phone is in use, a busy signal is returned and no antenna or subscriber equipment is involved at all in the call.

When the subscriber's phone is not in use, the MTCO will attempt to locate the subscriber (or, more specifically, locate the antenna nearest the subscriber). This is done by broadcasting the subscriber's identification number and listening for a response. The antenna's receiving responses report the power level of the response to the MTCO, which determines the closest antenna to the subscriber. If the subscriber is out of range or inoperable, the call will time out and all FSMs will return to the initial < 0, 0, 0 > state. A subscriber who answers an incoming call by taking the phone off hook causes the connection to be made and the FSMs transition to state < 5, 12, 5 >.

From the connected state, < 5, 12, 5 >, figure 26 can again be used to show how a call can either terminate or the mobile subscriber can move out of the servicing antenna's cell (resulting in the call being handed off to another cell's antenna).



(SEE FIGURE 26 FOR CONTINUATION)

FIGURE 27. SUBSCRIBER-DESTINATION CALL ANALYSIS

C. SUMMARY OF CELLULAR RADIO PROTOCOL ANALYSIS

The preceding protocol analysis has shown one possible protocol which could be used in cellular radio operations. While several simplifying assumptions were made to make the illustration easier to follow, the analysis has proven the reachability of each of the possible FSM states for the three cellular radio components during the course of an outgoing or incoming call.

VI. CONCLUSION

Radio built upon the technology of the telegraph and telephone to change the way the world communicates. It was research in the field of radio that brought us such important electronic advances as the development of the triode (amplification), the antenna, and many of the foundations upon which the television was built.

Numerous inventors were involved in the development of the technology. With the arguable exception of Gugliemo Marconi, no one inventor can be credited with more than a small proportion of the components that make up the earliest, or more modern, radios. More than anyone else, it was Marconi who was responsible for the explosion of interest in radio by demonstrating radio's potential and, subsequently, commercializing it.

Television, the most important mass media tool of modern times, borrowed heavily from developments in the radio industry for its technology, standardization, and regulations. While the "wireless" aspect of television was not the driving factor behind the desire to be able to transmit pictures, the advantages of using electromagnetic radiation to send television signals was recognized early on. Television has gone through three significant generations of development (mechanical, electronic, and color) and has just begun its forth--High Definition Television (HDTV).

HDTV marks a turning point in the direction and speed of technological advances in this field. Today, the rate that technology is capable of advancing exceeds the rate at which we can standardize it. For this reason, government involvement in regulation and standard setting is playing a major (pre-commercialization) role in the pace of development for this "next generation" of television.

Other areas of science, to include computer science, have benefited greatly from the contributions of both radio and television-related research. The earliest of computers can trace the origin of their components to the vacuum tubes developed for the radio and the cathode ray tubes used for television. Today's computers rely on other aspects of radio research in the areas of networking and communication. To summarize, the technologies, from an historical aspect, can trace their common roots to those of the radio.

To illustrate a modern-day application of the technology, this paper has shown how radio has been applied to the mobile telephone industry. Borrowing from tools used in computer science, a reachability analysis for a possible cellular radio protocol has been illustrated using finite state machines

LIST OF REFERENCES

[AITK76] Aitken, Hugh G. J., *Syntony and Spark-The Origins of Radio*, John Wiley & Sons, New York, 1976.

[ARCH38] Archer, Gleason L., *History of Radio to 1926*, American Book-Stratford Press, New York, 1938.

[BALA92] Balanis, Constantine A., Antenna Theory: A Review, *Proceedings of the IEEE*, VOL. 80, No. 1, January 1992.

[BELT91] Beltz, Cynthia A., *High-Tech Maneuvers: Industiral Policy Lessons of HDTV*, AEI Press, Washington, D. C., 1991.

[DAFM83] Department of the Army, "Communications-Electronics Fundamentals: Electronic Tube Theory and Circuits," U.S. Government Printing Office, September 1983.

[DAFM85] Department of the Army, "Communications-Electronics Fundamentals: Transmission Lines, Wave Propagation, and Antennas" U.S. Government Printing Office, June1985.

[DUNL44] Dunlap, Orrin E., *Radio's 100 Men of Science*, Harper & Brothers, New York, 1944.

[FREE58] Freebody, J. W., *Telegraphy*, Sir Isaac Pitman & Sons, Ltd., London, 1958.

[FREE91] Freeman, Roger L., *Telecommunication Transmission Handbook*, John Wiley & Sons, New York, 1991.

[MACL49] Maclaurin, W. Rupert, *Invention & Innovation in the Radio Industry*, Macmillan Company, New York, 1949.

[RYAN90] Ryan, Bob, "Farewell to Chips?", *BYTE Magazine*, January 1990, pp.237-248

[SCRO71] Scroggie, Maurice G., *Foundations of Wireless and Electronics*, Butterworth & Co., London, 1971.

[SHIE77] Shiers, George, *Technical Development of Television*, Arno Press, New York, 1977.

INITIAL DISTRIBUTION LIST

Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
Dudley Knox Library Code 052 Naval Postgraduate School Monterey, CA 93943	2
Chairman, Code CS Computer Science Department Naval Postgraduate School Monterey, CA 93943	2
CPT(P) William D. Reese CCSA-E CMR 405 Box 3062 APO, AE 09034	1