

Radio Electronics and Engineering

Dean Shell

First Edition, 2012

ISBN 978-81-323-3212-1

© All rights reserved.

Published by:

Research World

4735/22 Prakashdeep Bldg,

Ansari Road, Darya Ganj,

Delhi - 110002

Email: info@wtbooks.com

Table of Contents

Chapter 1 - RF Connector

Chapter 2 - Electronic Filter

Chapter 3 - Transmission Line

Chapter 4 - Voltage-Controlled Oscillator

Chapter 5 - Electronic Amplifier

Chapter 6 - Transmitter

Chapter 7 - Antenna

Chapter 8 - Receiver (Radio)

Chapter 9 - Detector

Chapter 10 - Electromagnetic Shielding

Chapter 11 - Radio-Frequency Identification

Chapter 12 - Amateur Radio Repeater

Chapter 13 - Magnetic Detector and Radio Receiver Design

Chapter 1

RF Connector



Male Type N RF connector.

A coaxial **RF connector** is an electrical connector designed to work at radio frequencies in the multi-megahertz range. RF connectors are typically used with coaxial cables and are designed to maintain the shielding that the coaxial design offers. Better models also minimize the change in transmission line impedance at the connection. Mechanically they provide a fastening mechanism (thread, bayonet, braces, push pull) and springs for a low

ohmic electric contact while sparing the gold surface thus allowing above 1000 reconnects and reducing the insertion force. Research activity in the area of radio-frequency (RF) circuit design has surged in the last decade in direct response to the enormous market demand for inexpensive, high data rate wireless transceivers.

Types



The coaxial connector **Miniquick**



N right angle direct male connector, solder-type, for semi-rigid .141 cable

Standard types

- 7/16 DIN connector, a high power 50 Ω connector originally developed by Spinner

- BNC (bayonet Neill-Concelman)
- C connector (Concelman)
- Dezifix connector, hermaphrodite connector used mainly by Rohde & Schwarz
- IEC 169-2 connector, also called *Belling Lee connector* or *PAL*, used throughout Europe and some other countries for domestic television installations and as *FM connector* for radio. It is standardised in *EN 60169-2*.
- F connector, used for domestic television installations and domestic satellite LNBs (75 Ω) world wide.
- GR connector (General Radio)
- HN connector, a high voltage version of the N connector
- MiniQuick
- Motorola connector, standard AM/FM antenna connector used for automotive radios
- Musa connector, a 50 Ω connector used in telecommunications and broadcast video
- N connector (Neill)
- NMO mount (new Motorola mount), for removable mobile antennas. Large threaded base for durability in wind.
- SC connector, screw version of C connector
- TNC connector (threaded Neill-Concelman)
- Twin lead
- UHF connector (e.g., PL-259/SO-239). Also referred to as an *M-type* connector by Japanese manufacturers such as Kenwood

Miniature types

- Miniature BNC connectors
- Miniature UHF connectors
- DIN 47223 connectors
- IPX connector
- SMZ connector - System 43 (BT43 and High Density HD43) for use in DDF

Micro-miniature types

- IMP connector
- MMT connector
- MMS connector
- U.FL connector
- UMP connector

Sub-miniature types

- MMCX connector
- MCX connector
- FME connector
- SMA connector, including variants:

- 3.5 and 2.92 mm connectors, which cross-mate with SMA, and
 - 2.4, 1.85 and 1.0 mm connectors, which do not cross-mate with SMA
- SMB connector
- SMC connector
- SMP connector

Precision types

- APC-7 connector

Flange connectors

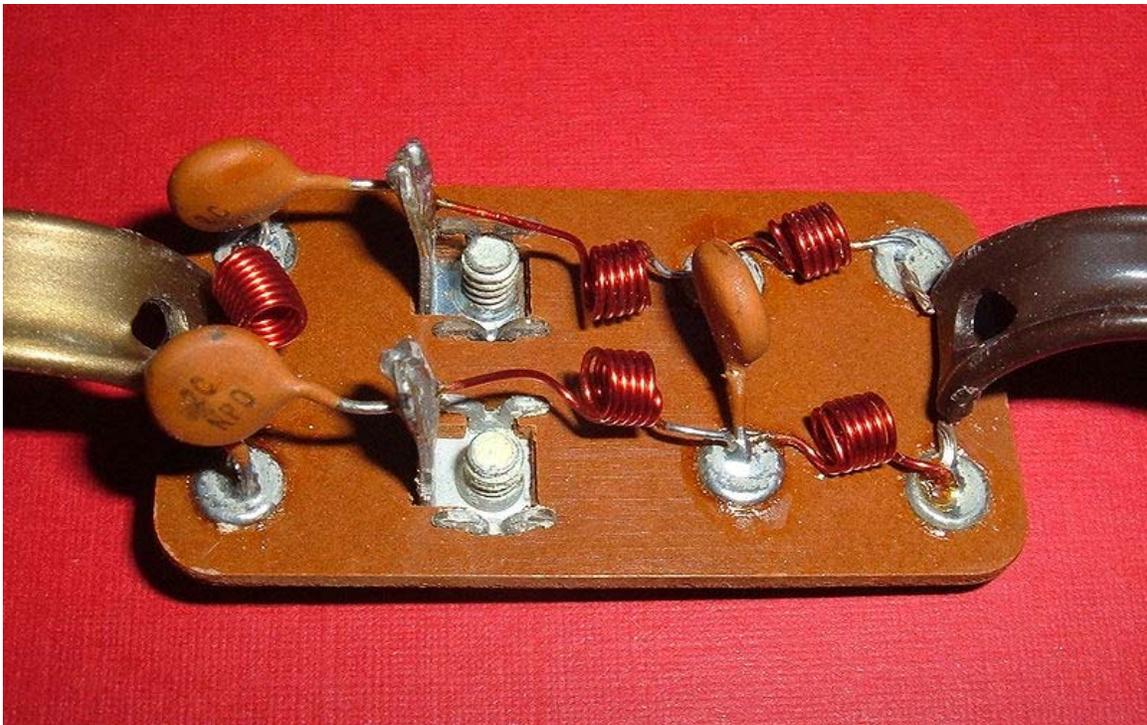
- EIA RF Connectors series of RF flange connectors

Quick-lock connectors

- QMA and QN connector
- QLS connector
- SnapN connector
- Mini-QMA connector
- WQMA (Waterproof QMA)
- HPQN connector

Chapter 2

Electronic Filter



Television signal splitter consisting of a high-pass filter (left) and a low-pass filter (right). The antenna is connected to the screw terminals to the left of center.

Electronic filters are electronic circuits which perform signal processing functions, specifically to remove unwanted frequency components from the signal, to enhance wanted ones, or both. Electronic filters can be:

- passive or active
- analog or digital
- high-pass, low-pass, bandpass, band-reject (band reject; notch), or all-pass.
- discrete-time (sampled) or continuous-time

- linear or non-linear
- infinite impulse response (IIR type) or finite impulse response (FIR type)

The most common types of electronic filters are linear filters, regardless of other aspects of their design.

History

The oldest forms of electronic filters are passive analog linear filters, constructed using only resistors and capacitors or resistors and inductors. These are known as RC and RL single-pole filters respectively. More complex multipole LC filters have also existed for many years, and their operation is well understood.

Hybrid filters are also possible, typically involving a combination of analog amplifiers with mechanical resonators or delay lines. Other devices such as CCD delay lines have also been used as discrete-time filters. With the availability of digital signal processing, active digital filters have become common.

Classification by technology

Passive filters

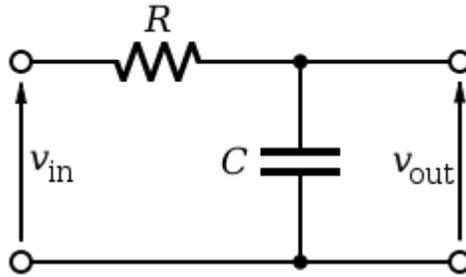
Passive implementations of linear filters are based on combinations of resistors (R), inductors (L) and capacitors (C). These types are collectively known as *passive filters*, because they do not depend upon an external power supply and/or they do not contain active components such as transistors.

Inductors block high-frequency signals and conduct low-frequency signals, while capacitors do the reverse. A filter in which the signal passes through an inductor, or in which a capacitor provides a path to ground, presents less attenuation to low-frequency signals than high-frequency signals and is a *low-pass filter*. If the signal passes through a capacitor, or has a path to ground through an inductor, then the filter presents less attenuation to high-frequency signals than low-frequency signals and is a *high-pass filter*. Resistors on their own have no frequency-selective properties, but are added to inductors and capacitors to determine the *time-constants* of the circuit, and therefore the frequencies to which it responds.

The inductors and capacitors are the reactive elements of the filter. The number of elements determines the order of the filter. In this context, an LC tuned circuit being used in a band-pass or band-stop filter is considered a single element even though it consists of two components.

At high frequencies (above about 100 megahertz), sometimes the inductors consist of single loops or strips of sheet metal, and the capacitors consist of adjacent strips of metal. These inductive or capacitive pieces of metal are called stubs.

Single element types



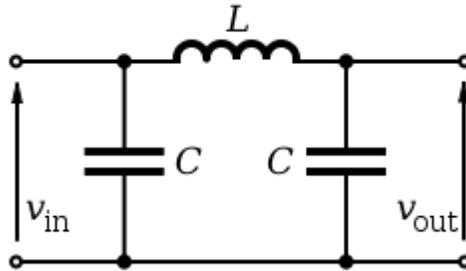
A low-pass electronic filter realised by an RC circuit

The simplest passive filters, RC and RL filters, include only one reactive element, except hybrid LC filter which is characterized by inductance and capacitance integrated in one element.

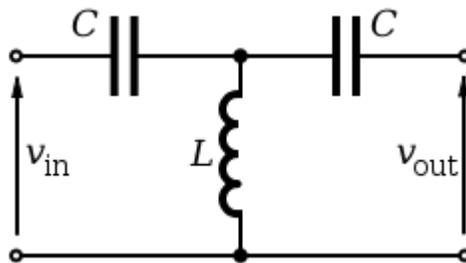
L filter

An L filter consists of two reactive elements, one in series and one in parallel.

T and π filters



Low-pass π filter



High-pass T filter

Three-element filters can have a 'T' or ' π ' topology and in either geometries, a low-pass, high-pass, band-pass, or band-stop characteristic is possible. The components can be chosen symmetric or not, depending on the required frequency characteristics. The high-pass T filter in the illustration, has a very low impedance at high frequencies, and a very

high impedance at low frequencies. That means that it can be inserted in a transmission line, resulting in the high frequencies being passed and low frequencies being reflected. Likewise, for the illustrated low-pass π filter, the circuit can be connected to a transmission line, transmitting low frequencies and reflecting high frequencies. Using m -derived filter sections with correct termination impedances, the input impedance can be reasonably constant in the pass band.

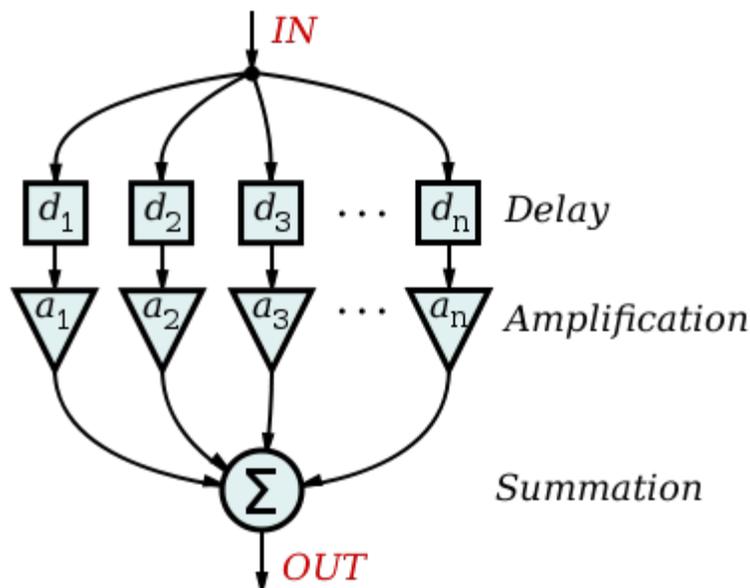
Multiple element types

Multiple element filters are usually constructed as a ladder network. These can be seen as a continuation of the L,T and π designs of filters. More elements are needed when it is desired to improve some parameter of the filter such as stop-band rejection or slope of transition from pass-band to stop-band.

Active filters

Active filters are implemented using a combination of passive and active (amplifying) components, and require an outside power source. Operational amplifiers are frequently used in active filter designs. These can have high Q factor, and can achieve resonance without the use of inductors. However, their upper frequency limit is limited by the bandwidth of the amplifiers used.

Digital filters



A general finite impulse response filter with n stages, each with an independent delay, d_i and amplification gain, a_i .

Digital signal processing allows the inexpensive construction of a wide variety of filters. The signal is sampled and an analog-to-digital converter turns the signal into a stream of numbers. A computer program running on a CPU or a specialized DSP (or less often

running on a hardware implementation of the algorithm) calculates an output number stream. This output can be converted to a signal by passing it through a digital-to-analog converter. There are problems with noise introduced by the conversions, but these can be controlled and limited for many useful filters. Due to the sampling involved, the input signal must be of limited frequency content or aliasing will occur.

Other filter technologies

Quartz filters and piezoelectrics

In the late 1930s, engineers realized that small mechanical systems made of rigid materials such as quartz would acoustically resonate at radio frequencies, i.e. from audible frequencies (sound) up to several hundred megahertz. Some early resonators were made of steel, but quartz quickly became favored. The biggest advantage of quartz is that it is piezoelectric. This means that quartz resonators can directly convert their own mechanical motion into electrical signals. Quartz also has a very low coefficient of thermal expansion which means that quartz resonators can produce stable frequencies over a wide temperature range. Quartz crystal filters have much higher quality factors than LCR filters. When higher stabilities are required, the crystals and their driving circuits may be mounted in a "crystal oven" to control the temperature. For very narrow band filters, sometimes several crystals are operated in series.

Engineers realized that a large number of crystals could be collapsed into a single component, by mounting comb-shaped evaporations of metal on a quartz crystal. In this scheme, a "tapped delay line" reinforces the desired frequencies as the sound waves flow across the surface of the quartz crystal. The tapped delay line has become a general scheme of making high- Q filters in many different ways.

SAW filters

SAW (surface acoustic wave) filters are electromechanical devices commonly used in radio frequency applications. Electrical signals are converted to a mechanical wave in a device constructed of a piezoelectric crystal or ceramic; this wave is delayed as it propagates across the device, before being converted back to an electrical signal by further electrodes. The delayed outputs are recombined to produce a direct analog implementation of a finite impulse response filter. This hybrid filtering technique is also found in an analog sampled filter. SAW filters are limited to frequencies up to 3 GHz.

BAW filters

BAW (Bulk Acoustic Wave) filters are electromechanical devices. BAW filters can implement ladder or lattice filters. BAW filters typically operate at frequencies from around 2 to around 16 GHz, and may be smaller or thinner than equivalent SAW filters. Two main variants of BAW filters are making their way into devices, Thin film bulk acoustic resonator or FBAR and Solid Mounted Bulk Acoustic Resonators.

Garnet filters

Another method of filtering, at microwave frequencies from 800 MHz to about 5 GHz, is to use a synthetic single crystal yttrium iron garnet sphere made of a chemical combination of yttrium and iron (**YIGF**, or **yttrium iron garnet filter**). The garnet sits on a strip of metal driven by a transistor, and a small loop antenna touches the top of the sphere. An electromagnet changes the frequency that the garnet will pass. The advantage of this method is that the garnet can be tuned over a very wide frequency by varying the strength of the magnetic field.

Atomic filters

For even higher frequencies and greater precision, the vibrations of atoms must be used. Atomic clocks use caesium masers as ultra-high Q filters to stabilize their primary oscillators. Another method, used at high, fixed frequencies with very weak radio signals, is to use a ruby maser tapped delay line.

The transfer function

The transfer function $H(s)$ of a filter is the ratio of the output signal $Y(s)$ to that of the input signal $X(s)$ as a function of the complex frequency s :

$$H(s) = \frac{Y(s)}{X(s)}$$

with $s = \sigma + j\omega$.

The transfer function of all linear time-invariant filters, when constructed of discrete components, will be the ratio of two polynomials in s , i.e. a rational function of s . The order of the transfer function will be the highest power of s encountered in either the numerator or the denominator.

Classification by topology

Electronic filters can be classified by the technology used to implement them. Filters using passive filter and active filter technology can be further classified by the particular electronic filter topology used to implement them.

Any given filter transfer function may be implemented in any electronic filter topology.

Some common circuit topologies are:

- Cauer topology - Passive
- Sallen Key topology - Active

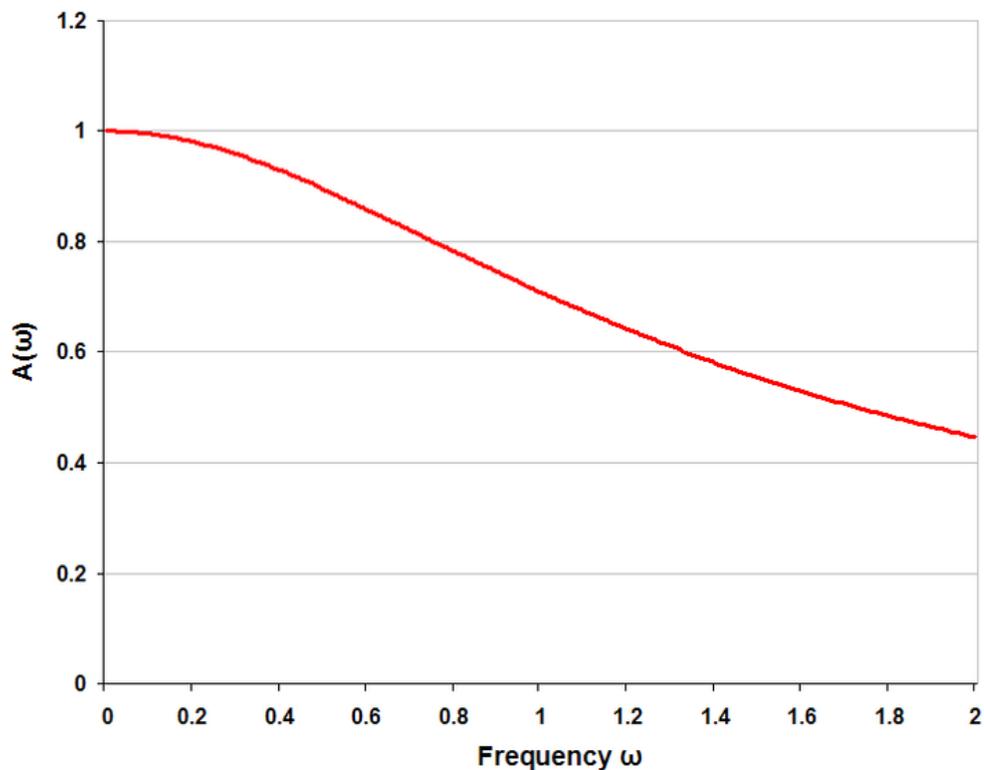
- Multiple Feedback topology - Active
- State Variable Topology - Active
- Biquadratic topology biquad filter - Active

Classification by design methodology

Historically, linear analog filter design has evolved through three major approaches. The oldest designs are simple circuits where the main design criterion was the Q factor of the circuit. This reflected the radio receiver application of filtering as Q was a measure of the frequency selectivity of a tuning circuit. From the 1920s filters began to be designed from the image point of view, mostly being driven by the requirements of telecommunications. After World War II the dominant methodology was network synthesis. The higher mathematics used originally required extensive tables of polynomial coefficient values to be published but modern computer resources have made that unnecessary.

Direct circuit analysis

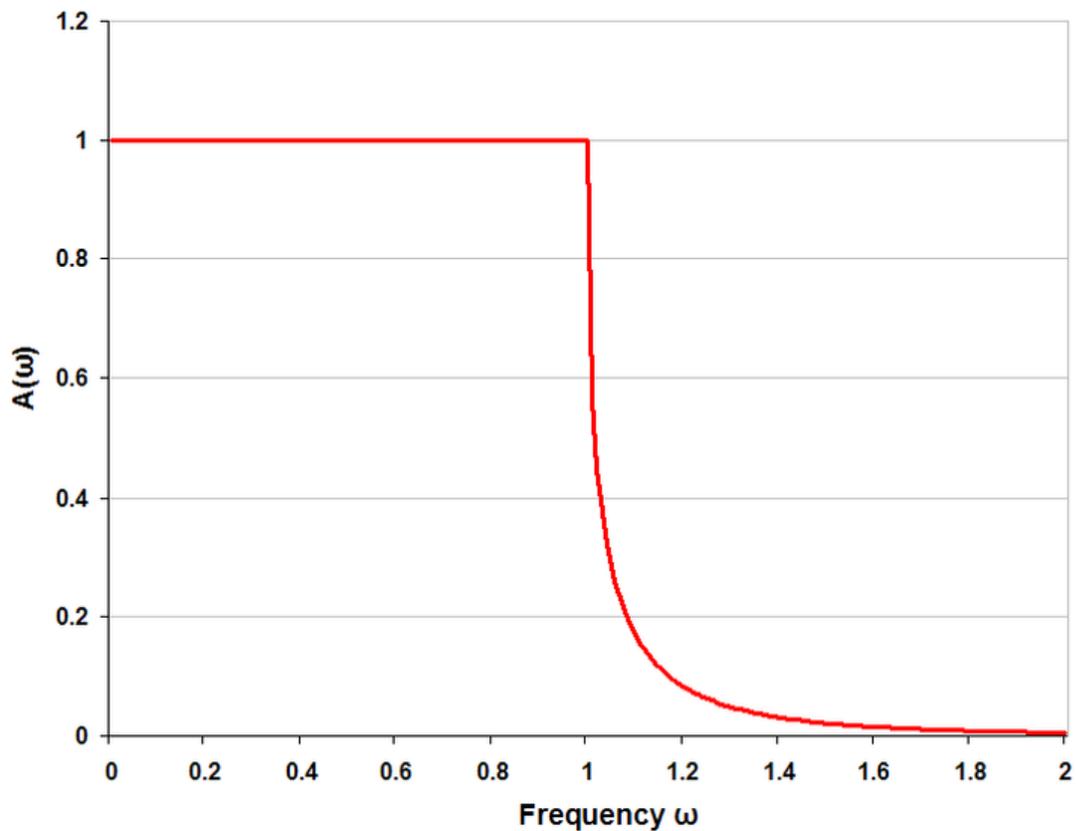
Low order filters can be designed by directly applying basic circuit laws such as Kirchoff's laws to obtain the transfer function. This kind of analysis is usually only carried out for simple filters of 1st or 2nd order.



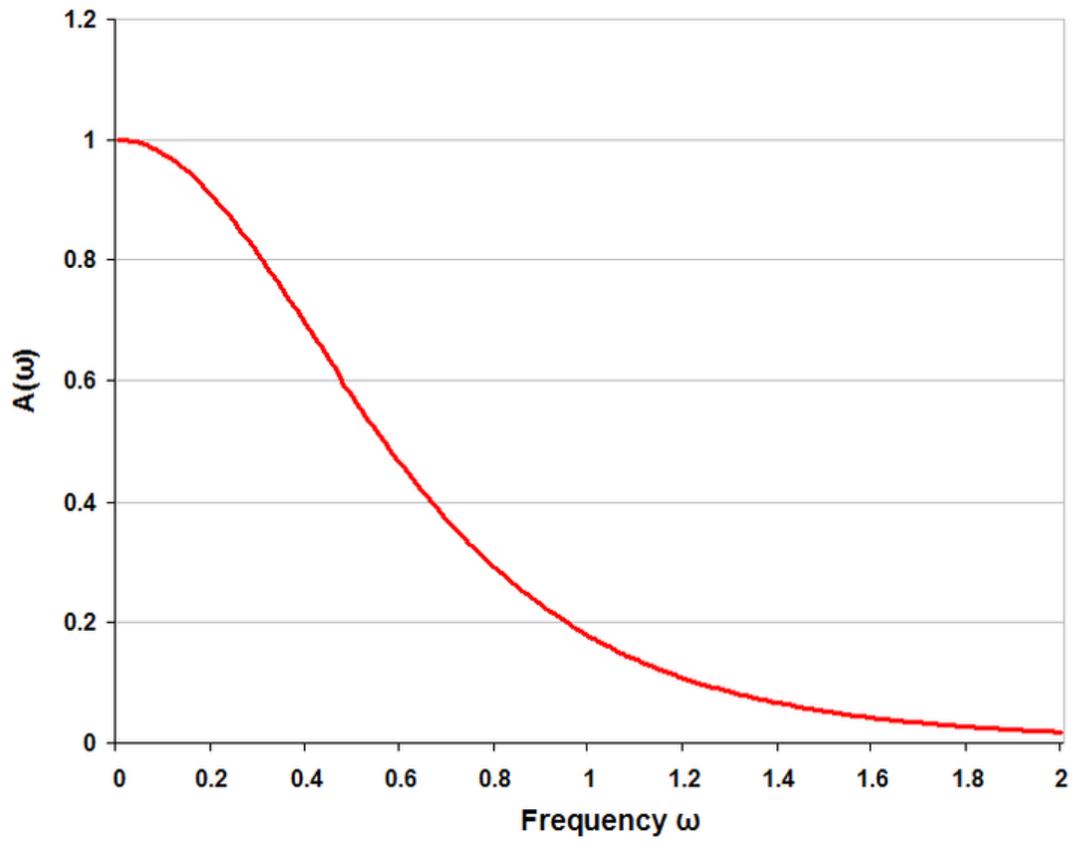
RL filter frequency response

Image impedance analysis

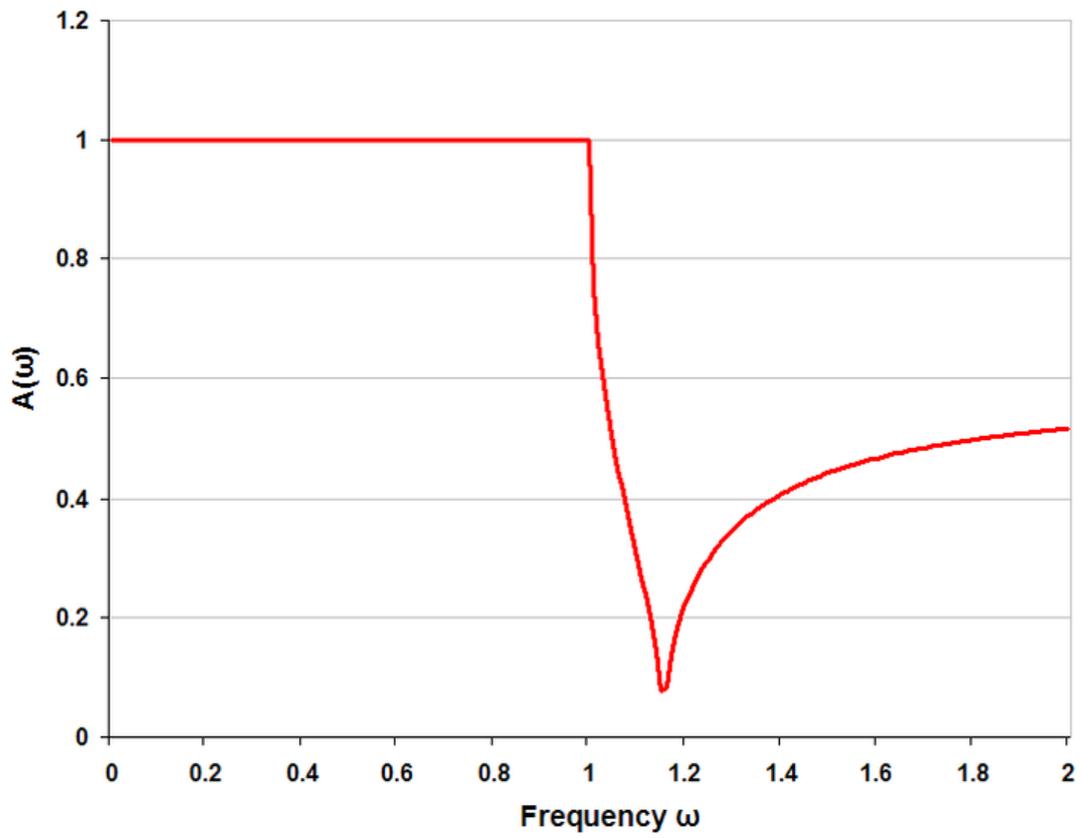
This approach analyses the filter sections from the point of view of the filter being in an infinite chain of identical sections. It has the advantages of simplicity of approach and the ability to easily extend to higher orders. It has the disadvantage that accuracy of predicted responses relies on filter terminations in the image impedance, which is usually not the case.



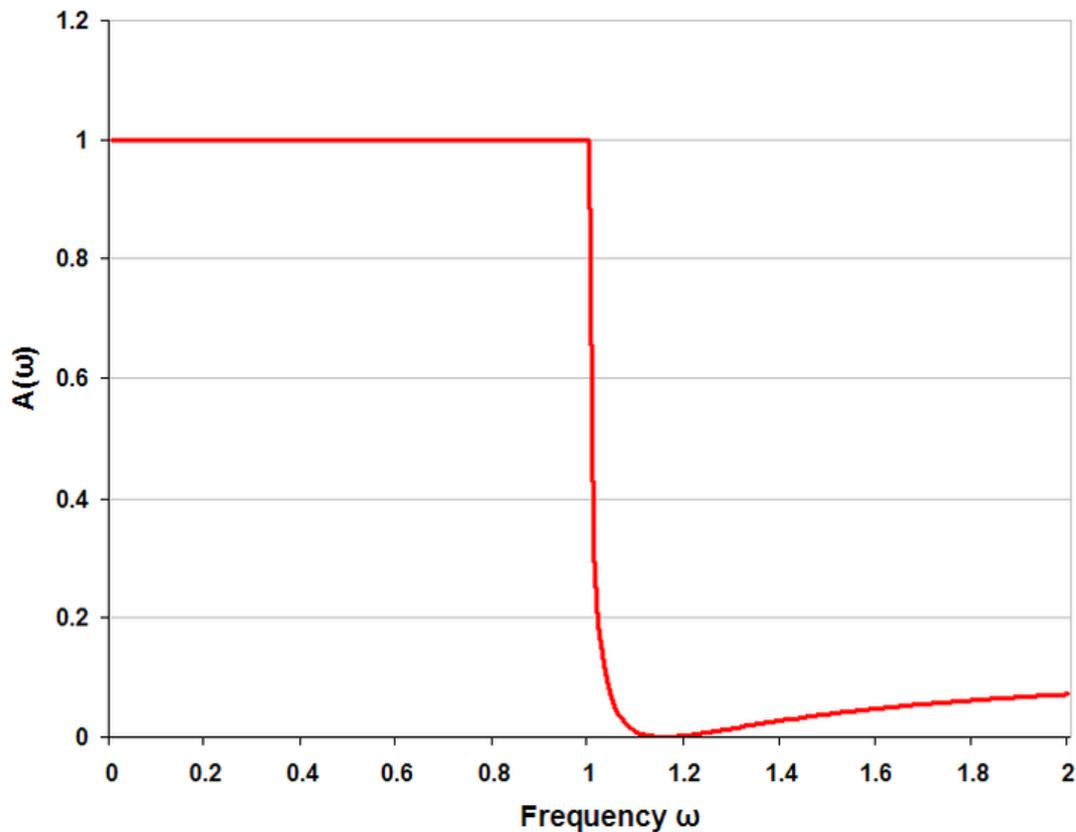
Constant k filter response with 5 elements



Zobel network (constant R) filter, 5 sections



m-derived filter response, $m=0.5$, 2 elements

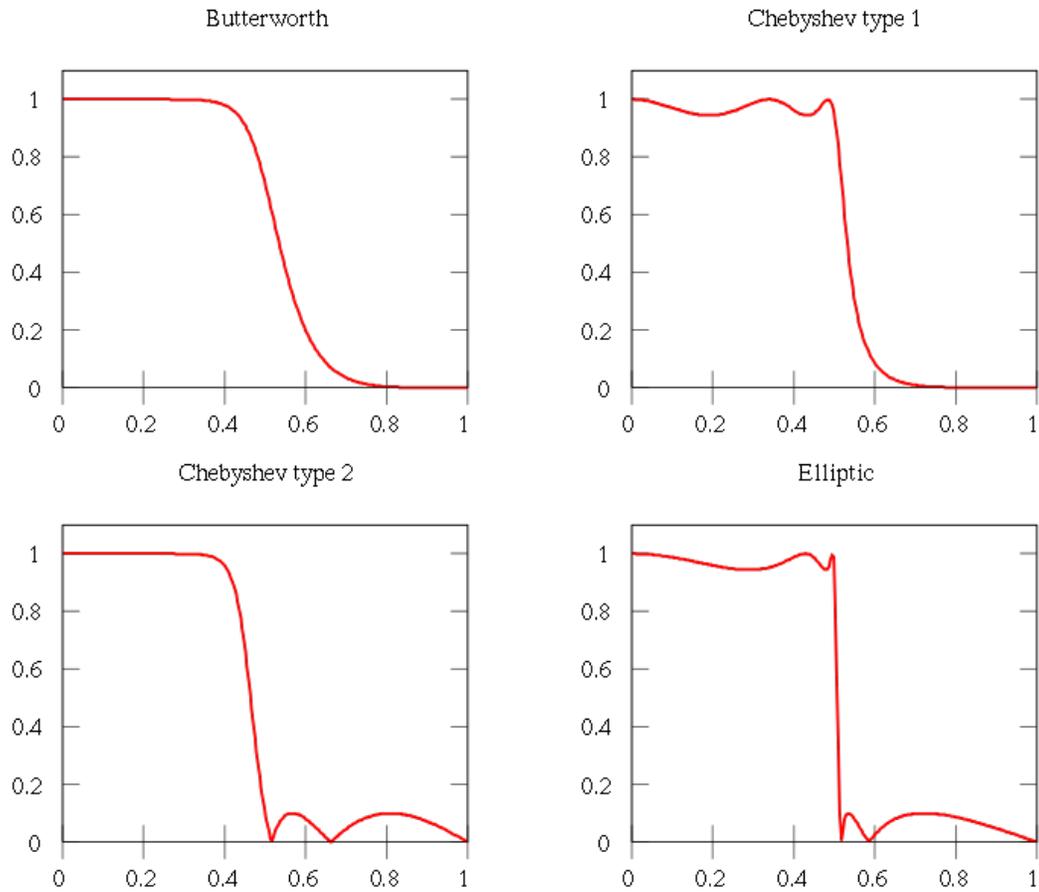


m-derived filter response, $m=0.5$, 5 elements

Network synthesis

The network synthesis approach starts with a required transfer function and then expresses that as a polynomial equation of the input impedance of the filter. The actual element values of the filter are obtained by continued-fraction or partial-fraction expansions of this polynomial. Unlike the image method, there is no need for impedance matching networks at the terminations as the effects of the terminating resistors are included in the analysis from the start.

Here is an image comparing Butterworth, Chebyshev, and elliptic filters. The filters in this illustration are all fifth-order low-pass filters. The particular implementation – analog or digital, passive or active – makes no difference; their output would be the same.



As is clear from the image, elliptic filters are sharper than all the others, but they show ripples on the whole bandwidth.

Chapter 3

Transmission Line

A **transmission line** is a material medium or structure that forms a path for directing the transmission of energy from one place to another, such as electromagnetic waves or acoustic waves, as well as electric power transmission.

However in communications and electronic engineering, the term has a more specific meaning. In these fields, transmission lines are specialized cables and other media designed to carry alternating current and electromagnetic waves of radio frequency, that is, currents with a frequency high enough that its wave nature must be taken into account. Transmission lines are used for purposes such as connecting radio transmitters and receivers with their antennas, distributing cable television signals, and computer network connections.

Ordinary electrical cables suffice to carry low frequency AC, such as mains power, which reverses direction 50 to 60 times per second. However, they cannot be used to carry currents in the radio frequency range or higher, which reverse direction millions to billions of times per second, because the energy tends to radiate off the cable as radio waves, causing power losses. Radio frequency currents also tend to reflect from discontinuities in the cable such as connectors, and travel back down the cable toward the source. These reflections act as bottlenecks, preventing the power from reaching the destination. Transmission lines use specialized construction such as precise conductor dimensions and spacing, and impedance matching, to carry electromagnetic signals with minimal reflections and power losses. Types of transmission line include ladder line, coaxial cable, dielectric slabs, stripline, optical fiber, and waveguides. The higher the frequency, the shorter are the waves in a transmission medium. Transmission lines must be used when the frequency is high enough that the wavelength of the waves begins to approach the length of the cable used. To conduct energy at frequencies above the radio range, such as millimeter waves, infrared, and light, the waves become much smaller than the dimensions of the structures used to guide them, so transmission line techniques become inadequate and the methods of optics are used.

The theory of sound wave propagation is very similar mathematically to that of electromagnetic waves, so techniques from transmission line theory are also used to build structures to conduct acoustic waves; and these are also called transmission lines.

History

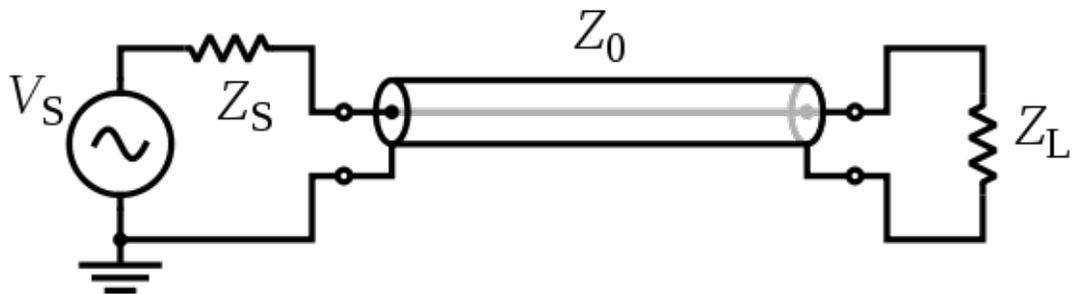
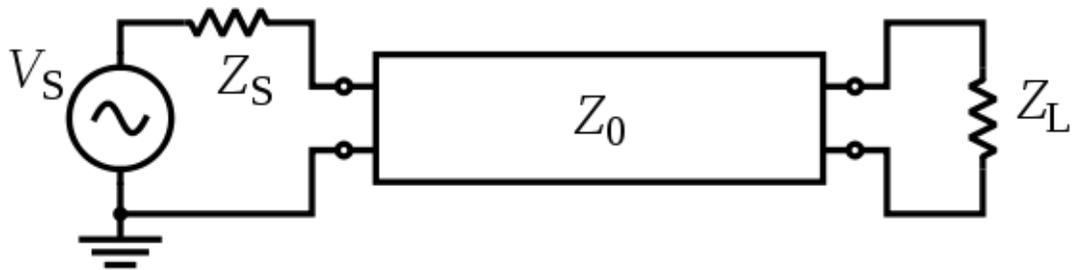
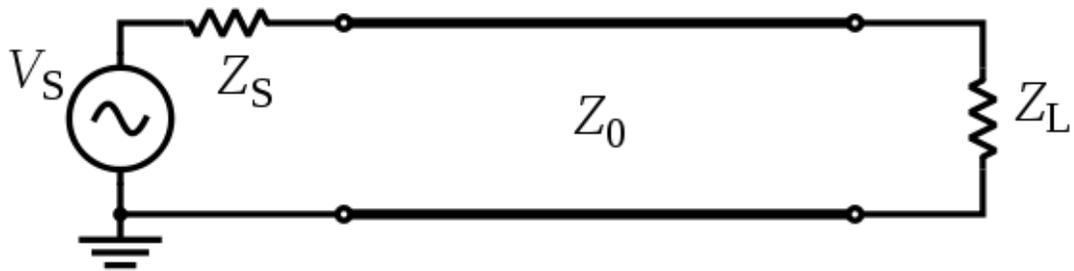
Mathematical analysis of the behaviour of electrical transmission lines grew out of the work of James Clerk Maxwell, Lord Kelvin and Oliver Heaviside. In 1855 Lord Kelvin formulated a diffusion model of the current in a submarine cable. The model correctly predicted the poor performance of the 1858 trans-Atlantic submarine telegraph cable. In 1885 Heaviside published the first papers that described his analysis of propagation in cables and the modern form of the telegrapher's equations.

Applicability

In many electric circuits, the length of the wires connecting the components can for the most part be ignored. That is, the voltage on the wire at a given time can be assumed to be the same at all points. However, when the voltage changes in a time interval comparable to the time it takes for the signal to travel down the wire, the length becomes important and the wire must be treated as a transmission line. Stated another way, the length of the wire is important when the signal includes frequency components with corresponding wavelengths comparable to or less than the length of the wire.

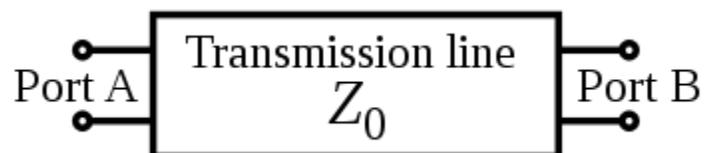
A common rule of thumb is that the cable or wire should be treated as a transmission line if the length is greater than $1/10$ of the wavelength. At this length the phase delay and the interference of any reflections on the line become important and can lead to unpredictable behavior in systems which have not been carefully designed using transmission line theory.

The four terminal model



Variations on the schematic electronic symbol for a transmission line.

For the purposes of analysis, an electrical transmission line can be modelled as a two-port network (also called a quadripole network), as follows:



In the simplest case, the network is assumed to be linear (i.e. the complex voltage across either port is proportional to the complex current flowing into it when there are no reflections), and the two ports are assumed to be interchangeable. If the transmission line is uniform along its length, then its behaviour is largely described by a single parameter called the *characteristic impedance*, symbol Z_0 . This is the ratio of the complex voltage of a given wave to the complex current of the same wave at any point on the line. Typical values of Z_0 are 50 or 75 ohms for a coaxial cable, about 100 ohms for a twisted pair of wires, and about 300 ohms for a common type of untwisted pair used in radio transmission.

When sending power down a transmission line, it is usually desirable that as much power as possible will be absorbed by the load and as little as possible will be reflected back to the source. This can be ensured by making the load impedance equal to Z_0 , in which case the transmission line is said to be *matched*.

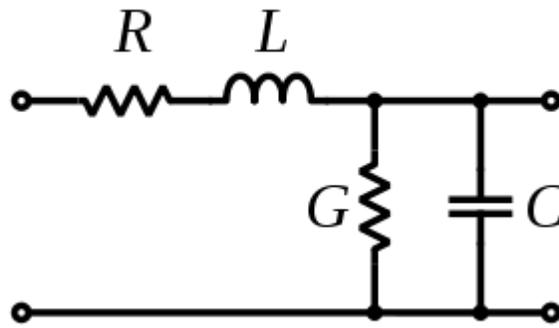
Some of the power that is fed into a transmission line is lost because of its resistance. This effect is called *ohmic* or *resistive* loss. At high frequencies, another effect called *dielectric loss* becomes significant, adding to the losses caused by resistance. Dielectric loss is caused when the insulating material inside the transmission line absorbs energy from the alternating electric field and converts it to heat. The Transmission Line is modeled with a Resistance(R) and Inductance(L) in Series with a Capacitance(C) and Conductance(G) in Parallel. The Resistance and Conductance contributes to the loss of the Transmission Line.

The total loss of power in a transmission line is often specified in decibels per metre (dB/m), and usually depends on the frequency of the signal. The manufacturer often supplies a chart showing the loss in dB/m at a range of frequencies. A loss of 3 dB corresponds approximately to a halving of the power.

High-frequency transmission lines can be defined as those designed to carry electromagnetic waves whose wavelengths are shorter than or comparable to the length of the line. Under these conditions, the approximations useful for calculations at lower frequencies are no longer accurate. This often occurs with radio, microwave and optical signals, metal mesh optical filters, and with the signals found in high-speed digital circuits.

Telegrapher's equations

The **Telegrapher's Equations** (or just **Telegraph Equations**) are a pair of linear differential equations which describe the voltage and current on an electrical transmission line with distance and time. They were developed by Oliver Heaviside who created the *transmission line model*, and are based on Maxwell's Equations.



Schematic representation of the elementary component of a transmission line.

The transmission line model represents the transmission line as an infinite series of two-port elementary components, each representing an infinitesimally short segment of the transmission line:

- The distributed resistance R of the conductors is represented by a series resistor (expressed in ohms per unit length).
- The distributed inductance L (due to the magnetic field around the wires, self-inductance, etc.) is represented by a series inductor (henries per unit length).
- The capacitance C between the two conductors is represented by a shunt capacitor C (farads per unit length).
- The conductance G of the dielectric material separating the two conductors is represented by a shunt resistor between the signal wire and the return wire (siemens per unit length).

The model consists of an *infinite series* of the elements shown in the figure, and that the values of the components are specified *per unit length* so the picture of the component can be misleading. R , L , C , and G may also be functions of frequency. An alternative notation is to use R' , L' , C' and G' to emphasize that the values are derivatives with respect to length. These quantities can also be known as the primary line constants to distinguish from the secondary line constants derived from them, these being the propagation constant, attenuation constant and phase constant.

The line voltage $V(x)$ and the current $I(x)$ can be expressed in the frequency domain as

$$\frac{\partial V(x)}{\partial x} = -(R + j\omega L)I(x)$$

$$\frac{\partial I(x)}{\partial x} = -(G + j\omega C)V(x)$$

When the elements R and G are negligibly small the transmission line is considered as a lossless structure. In this hypothetical case, the model depends only on the L and C

elements which greatly simplifies the analysis. For a lossless transmission line, the second order steady-state Telegrapher's equations are:

$$\frac{\partial^2 V(x)}{\partial x^2} + \omega^2 LC \cdot V(x) = 0$$

$$\frac{\partial^2 I(x)}{\partial x^2} + \omega^2 LC \cdot I(x) = 0$$

These are wave equations which have plane waves with equal propagation speed in the forward and reverse directions as solutions. The physical significance of this is that electromagnetic waves propagate down transmission lines and in general, there is a reflected component that interferes with the original signal. These equations are fundamental to transmission line theory.

If R and G are not neglected, the Telegrapher's equations become:

$$\frac{\partial^2 V(x)}{\partial x^2} = \gamma^2 V(x)$$

$$\frac{\partial^2 I(x)}{\partial x^2} = \gamma^2 I(x)$$

where

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)}$$

and the characteristic impedance is:

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

The solutions for $V(x)$ and $I(x)$ are:

$$V(x) = V^+ e^{-\gamma x} + V^- e^{\gamma x}$$

$$I(x) = \frac{1}{Z_0} (V^+ e^{-\gamma x} - V^- e^{\gamma x})$$

The constants V^\pm and I^\pm must be determined from boundary conditions. For a voltage pulse $V_{in}(t)$, starting at $x = 0$ and moving in the positive x -direction, then the transmitted pulse $V_{out}(x, t)$ at position x can be obtained by computing the Fourier Transform, $\tilde{V}(\omega)$, of $V_{in}(t)$, attenuating each frequency component by $e^{-\text{Re}(\gamma)x}$, advancing its

phase by $-\text{Im}(\gamma)x$, and taking the inverse Fourier Transform. The real and imaginary parts of γ can be computed as

$$\begin{aligned}\text{Re}(\gamma) &= (a^2 + b^2)^{1/4} \cos(\text{atan2}(b, a)/2) \\ \text{Im}(\gamma) &= (a^2 + b^2)^{1/4} \sin(\text{atan2}(b, a)/2)\end{aligned}$$

where atan2 is the two-parameter arctangent, and

$$\begin{aligned}a &\equiv \omega^2 LC \left[\left(\frac{R}{\omega L} \right) \left(\frac{G}{\omega C} \right) - 1 \right] \\ b &\equiv \omega^2 LC \left(\frac{R}{\omega L} + \frac{G}{\omega C} \right)\end{aligned}$$

For small losses and high frequencies, to first order in $R/\omega L$ and $G/\omega C$ one obtains

$$\begin{aligned}\text{Re}(\gamma) &\approx \frac{\sqrt{LC}}{2} \left(\frac{R}{L} + \frac{G}{C} \right) \\ \text{Im}(\gamma) &\approx \omega \sqrt{LC}\end{aligned}$$

Noting that an advance in phase by $-\omega\delta$ is equivalent to a time delay by δ , $V_{out}(t)$ can be simply computed as

$$V_{out}(x, t) \approx V_{in}(t - \sqrt{LC}x) e^{-\frac{\sqrt{LC}}{2} \left(\frac{R}{L} + \frac{G}{C} \right) x}$$

Input impedance of lossless transmission line

The characteristic impedance Z_0 of a transmission line is the ratio of the amplitude of a **single** voltage wave to its current wave. Since most transmission lines also have a reflected wave, the characteristic impedance is generally **not** the impedance that is measured on the line.

For a lossless transmission line, it can be shown that the impedance measured at a given position l from the load impedance Z_L is

$$Z_{in}(l) = Z_0 \frac{Z_L + jZ_0 \tan(\beta l)}{Z_0 + jZ_L \tan(\beta l)}$$

where $\beta = \frac{2\pi}{\lambda}$ is the wavenumber.

In calculating β , the wavelength is generally different inside the transmission line to what it would be in free-space and the velocity constant of the material the transmission line is made of needs to be taken into account when doing such a calculation.

Special cases

Half wave length

For the special case where $\beta l = n\pi$ where n is an integer (meaning that the length of the line is a multiple of half a wavelength), the expression reduces to the load impedance so that

$$Z_{\text{in}} = Z_L$$

for all n . This includes the case when $n = 0$, meaning that the length of the transmission line is negligibly small compared to the wavelength. The physical significance of this is that the transmission line can be ignored (i.e. treated as a wire) in either case.

Quarter wave length

For the case where the length of the line is one quarter wavelength long, or an odd multiple of a quarter wavelength long, the input impedance becomes

$$Z_{\text{in}} = \frac{Z_0^2}{Z_L}$$

Matched load

Another special case is when the load impedance is equal to the characteristic impedance of the line (i.e. the line is *matched*), in which case the impedance reduces to the characteristic impedance of the line so that

$$Z_{\text{in}} = Z_L = Z_0$$

for all l and all λ .

Short

For the case of a shorted load (i.e. $Z_L = 0$), the input impedance is purely imaginary and a periodic function of position and wavelength (frequency)

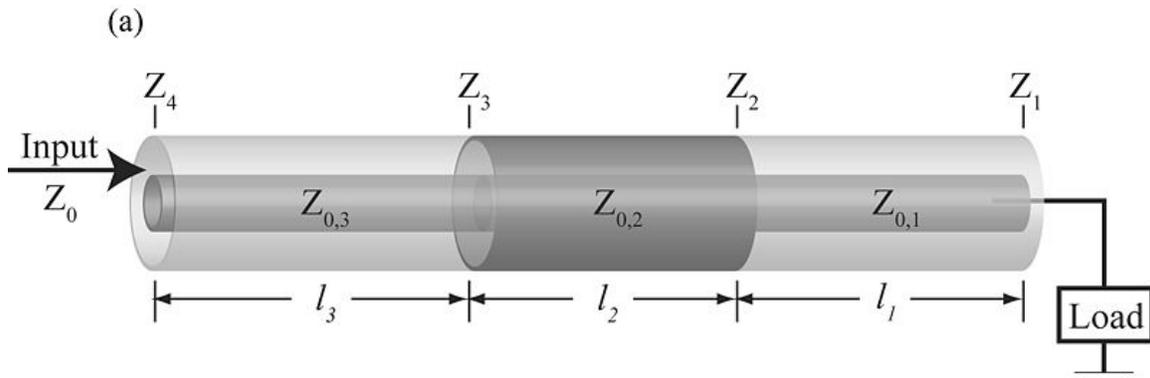
$$Z_{\text{in}}(l) = jZ_0 \tan(\beta l)$$

Open

For the case of an open load (i.e. $Z_L = \infty$), the input impedance is once again imaginary and periodic

$$Z_{in}(l) = -j Z_0 \cot(\beta l)$$

Stepped transmission line

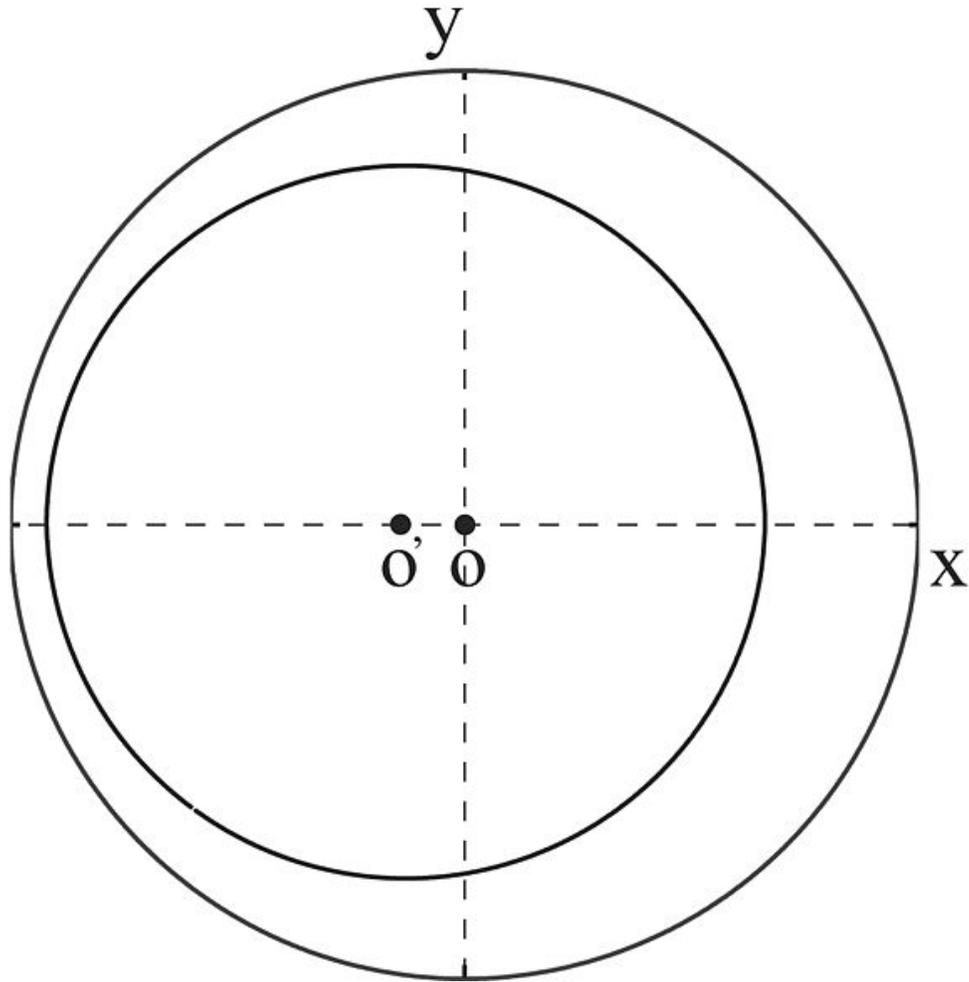


A simple example of stepped transmission line consisting of three segments.

Stepped transmission line is used for broad range impedance matching. It can be considered as multiple transmission line segments connected in serial, with the characteristic impedance of each individual element to be, $Z_{0,i}$. And the input impedance can be obtained from the successive application of the chain relation

$$Z_{i+1} = Z_{0,i} \frac{Z_i + j Z_{0,i} \tan(\beta_i l_i)}{Z_{0,i} + j Z_i \tan(\beta_i l_i)}$$

where β_i is the wave number of the i th transmission line segment and l_i is the length of this segment, and Z_i is the front-end impedance that loads the i th segment.



The impedance transformation circle along a transmission line whose characteristic impedance $Z_{0,i}$ is smaller than that of the input cable Z_0 . And as a result, the impedance curve is off-centered towards the -x axis. Conversely, if $Z_{0,i} > Z_0$, the impedance curve should be off-centered towards the +x axis.

Because the characteristic impedance of each transmission line segment $Z_{0,i}$ is often different from that of the input cable Z_0 , the impedance transformation circle is off centered along the x axis of the Smith Chart whose impedance representation is usually normalized against Z_0 .

Practical types

Coaxial cable

Coaxial lines confine virtually all of the electromagnetic wave to the area inside the cable. Coaxial lines can therefore be bent and twisted (subject to limits) without negative effects, and they can be strapped to conductive supports without inducing unwanted currents in them. In radio-frequency applications up to a few gigahertz, the wave

propagates in the transverse electric and magnetic mode (TEM) only, which means that the electric and magnetic fields are both perpendicular to the direction of propagation (the electric field is radial, and the magnetic field is circumferential). However, at frequencies for which the wavelength (in the dielectric) is significantly shorter than the circumference of the cable, transverse electric (TE) and transverse magnetic (TM) waveguide modes can also propagate. When more than one mode can exist, bends and other irregularities in the cable geometry can cause power to be transferred from one mode to another.

The most common use for coaxial cables is for television and other signals with bandwidth of multiple megahertz. In the middle 20th century they carried long distance telephone connections.

Microstrip

A microstrip circuit uses a thin flat conductor which is parallel to a ground plane. Microstrip can be made by having a strip of copper on one side of a printed circuit board (PCB) or ceramic substrate while the other side is a continuous ground plane. The width of the strip, the thickness of the insulating layer (PCB or ceramic) and the dielectric constant of the insulating layer determine the characteristic impedance. Microstrip is an open structure whereas coaxial cable is a closed structure.

Stripline

A stripline circuit uses a flat strip of metal which is sandwiched between two parallel ground planes. The insulating material of the substrate forms a dielectric. The width of the strip, the thickness of the substrate and the relative permittivity of the substrate determine the characteristic impedance of the strip which is a transmission line.

Balanced lines

A balanced line is a transmission line consisting of two conductors of the same type, and equal impedance to ground and other circuits. There are many formats of balanced lines, amongst the most common are twisted pair, star quad and twin-lead.

Twisted pair

Twisted pairs are commonly used for terrestrial telephone communications. In such cables, many pairs are grouped together in a single cable, from two to several thousand. The format is also used for data network distribution inside buildings, but in this case the cable used is more expensive with much tighter controlled parameters and either two or four pairs per cable.

Star quad

Star quad is another balanced format used at low frequencies. Applications include 4-wire telephony and microphone circuits.

Twin-lead

Twin-lead consists of a pair of conductors held apart by a continuous insulator.

Lecher lines

Lecher lines are a form of parallel conductor that can be used at UHF for creating resonant circuits. They are a convenient practical format that fills the gap between lumped components (used at HF/VHF) and resonant cavities (used at UHF/SHF).

Single-wire line

Unbalanced lines were formerly much used for telegraph transmission, but this form of communication has now fallen into disuse. Cables are similar to twisted pair in that many cores are bundled into the same cable but only one conductor is provided per circuit and there is no twisting. All the circuits on the same route use a common path for the return current (earth return). There is a power transmission version of single-wire earth return in use in many locations.

Waveguide

Waveguides are rectangular or circular metallic tubes inside which an electromagnetic wave is propagated and is confined by the tube. Waveguides are not capable of transmitting the transverse electromagnetic mode found in copper lines and must use some other mode. Consequently, they cannot be directly connected to cable and a mechanism for launching the waveguide mode must be provided at the interface.

Optical fiber

Optical fibers are a solid transparent fiber of glass or polymer which transmits a signal at optical, or near infrared, wavelengths. They form the backbone of all modern terrestrial communications networks due to the very high bandwidths that can be achieved. Optical fiber are another variety of waveguide.

General applications

Signal transfer

Electrical transmission lines are very widely used to transmit high frequency signals over long or short distances with minimum power loss. One familiar example is the down lead from a TV or radio aerial to the receiver.

Pulse generation

Transmission lines are also used as pulse generators. By charging the transmission line and then discharging it into a resistive load, a rectangular pulse equal in length to twice the electrical length of the line can be obtained, although with half the voltage. A Blumlein transmission line is a related pulse forming device that overcomes this limitation. These are sometimes used as the pulsed energy sources for radar transmitters and other devices.

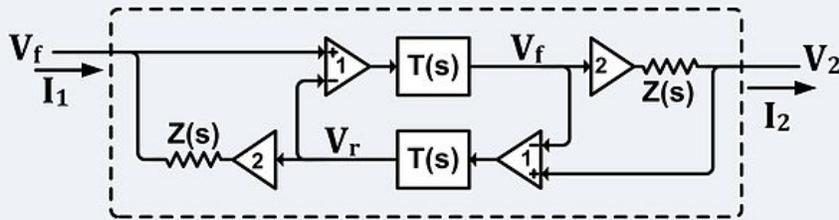
Stub filters

If a short-circuited or open-circuited transmission line is wired in parallel with a line used to transfer signals from point A to point B, then it will function as a filter. The method for making stubs is similar to the method for using Lecher lines for crude frequency measurement, but it is 'working backwards'. One method recommended in the RSGB's radiocommunication handbook is to take an open-circuited length of transmission line wired in parallel with the feeder delivering signals from an aerial. By cutting the free end of the transmission line, a minimum in the strength of the signal observed at a receiver can be found. At this stage the stub filter will reject this frequency and the odd harmonics, but if the free end of the stub is shorted then the stub will become a filter rejecting the even harmonics.

Acoustic transmission lines

An acoustic transmission line is the acoustic analog of the electrical transmission line, typically thought of as a rigid-walled tube that is long and thin relative to the wavelength of sound present in it.

Solutions of the Telegrapher's Equations as Circuit Components



$$\gamma(s) = \sqrt{(R_\omega + sL_\omega)(G_\omega + sC_\omega)}$$

$$Z(s) = \sqrt{(R_\omega + sL_\omega) / (G_\omega + sC_\omega)}$$

$$T(s) = e^{-X\gamma(s)}$$

where:

X = length of transmission line

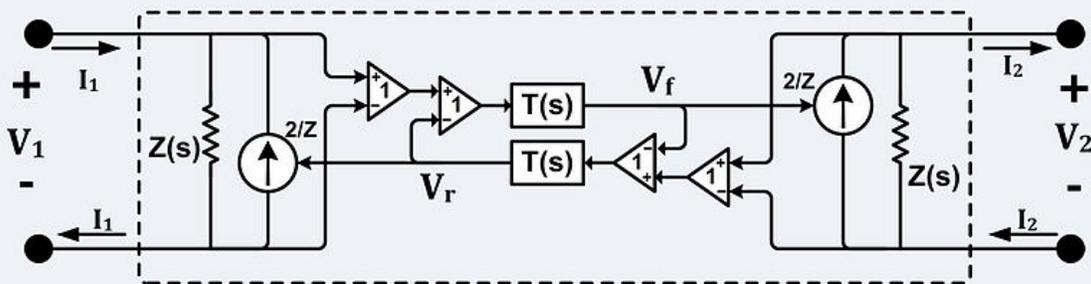
Z = characteristic impedance

T = propagation function

γ = propagation "constant"

Note: R_ω , L_ω , G_ω and C_ω may be functions of frequency

Equivalent circuit of a transmission line described by the Telegrapher's equations.



$2/Z$ = trans-admittance of VDCS (Voltage Dependent Current Source)

$$\gamma(s) = \sqrt{(R_\omega + sL_\omega)(G_\omega + sC_\omega)}$$

$$Z(s) = \sqrt{(R_\omega + sL_\omega) / (G_\omega + sC_\omega)}$$

$$T(s) = e^{-X\gamma(s)}$$

where:

X = length of transmission line

Z = characteristic impedance

T = propagation function

γ = propagation "constant"

$s = j\omega$

Note: R_ω , L_ω , G_ω and C_ω may be functions of frequency

Solutions of the Telegrapher's Equations as Components in the Equivalent Circuit of a Balanced Transmission Line Two-Port Implementation.

The solutions of the telegrapher's equations can be inserted directly into a circuit as components. The circuit in the left figure implements the solutions of the telegrapher's equations.

The right hand circuit is derived from the left hand circuit by source transformations. It also implements the solutions of the telegrapher's equations.

The solution of the telegrapher's equations can be expressed as an ABCD type *Two-port network* with the following defining equations

$$V_1 = V_2 \cosh(\gamma x) + I_2 Z \sinh(\gamma x)$$

$$I_1 = V_2 \frac{1}{Z} \sinh(\gamma x) + I_2 \cosh(\gamma x)$$

The symbols: E_s, E_L, I_s, I_L, l in the source book have been replaced by the symbols: V_1, V_2, I_1, I_2, x in the preceding two equations.

The ABCD type two-port gives V_1 and I_1 as functions of V_2 and I_2 . Both of the circuits above, when solved for V_1 and I_1 as functions of V_2 and I_2 yield exactly the same equations.

In the right hand circuit, all voltages except the port voltages are with respect to ground and the differential amplifiers have unshown connections to ground. An example of a transmission line modeled by this circuit would be a balanced transmission line such as a telephone line. The impedances $Z(s)$, the voltage dependent current sources (VDCSS) and the difference amplifiers (the triangle with the number "1") account for the interaction of the transmission line with the external circuit. The $T(s)$ blocks account for delay, attenuation, dispersion and whatever happens to the signal in transit. One of the $T(s)$ blocks carries the *forward wave* and the other carries the *backward wave*. The circuit, as depicted, is fully symmetric, although it is not drawn that way. The circuit depicted is equivalent to a transmission line connected from V_1 to V_2 in the sense that V_1, V_2, I_1 and I_2 would be same whether this circuit or an actual transmission line was connected between V_1 and V_2 . There is no implication that there are actually amplifiers inside the transmission line.

Every two-wire or balanced transmission line has an implicit (or in some cases explicit) third wire which may be called shield, sheath, common, Earth or ground. So every two-wire balanced transmission line has two modes which are nominally called the differential and common modes. The circuit shown on the right only models the differential mode.

In the left hand circuit, the voltage doublers, the difference amplifiers and impedances $Z(s)$ account for the interaction of the transmission line with the external circuit. This

circuit, as depicted, is also fully symmetric, and also not drawn that way. This circuit is a useful equivalent for an unbalanced transmission line like a coaxial cable or a micro strip line.

Chapter 4

Voltage-Controlled Oscillator

A **voltage-controlled oscillator** or **VCO** is an electronic oscillator designed to be controlled in oscillation frequency by a voltage input. The frequency of oscillation is varied by the applied DC voltage, while modulating signals may also be fed into the VCO to cause frequency modulation (FM) or phase modulation (PM); a VCO with digital pulse output may similarly have its repetition rate (FSK, PSK) or pulse width modulated (PWM).



A microwave (12-18 GHz) Voltage Controlled Oscillator

Types of VCOs

VCOs can be generally categorized into two groups based on the type of waveform produced: 1) harmonic oscillators, and 2) relaxation oscillators.

Harmonic oscillators generate a sinusoidal waveform. They consist of an amplifier that provides adequate gain and a resonant circuit that feeds back signal to the input. Oscillation occurs at the resonant frequency where a positive gain arises around the loop. Some examples of harmonic oscillators are crystal oscillators and LC-tank oscillators. When part of the resonant circuit's capacitance is provided by a varactor diode, the voltage applied to that diode varies the frequency.

Relaxation oscillators can generate a sawtooth or triangular waveform. They are commonly used in monolithic integrated circuits (ICs). They can provide a wide range of operational frequencies with a minimal number of external components. Relaxation

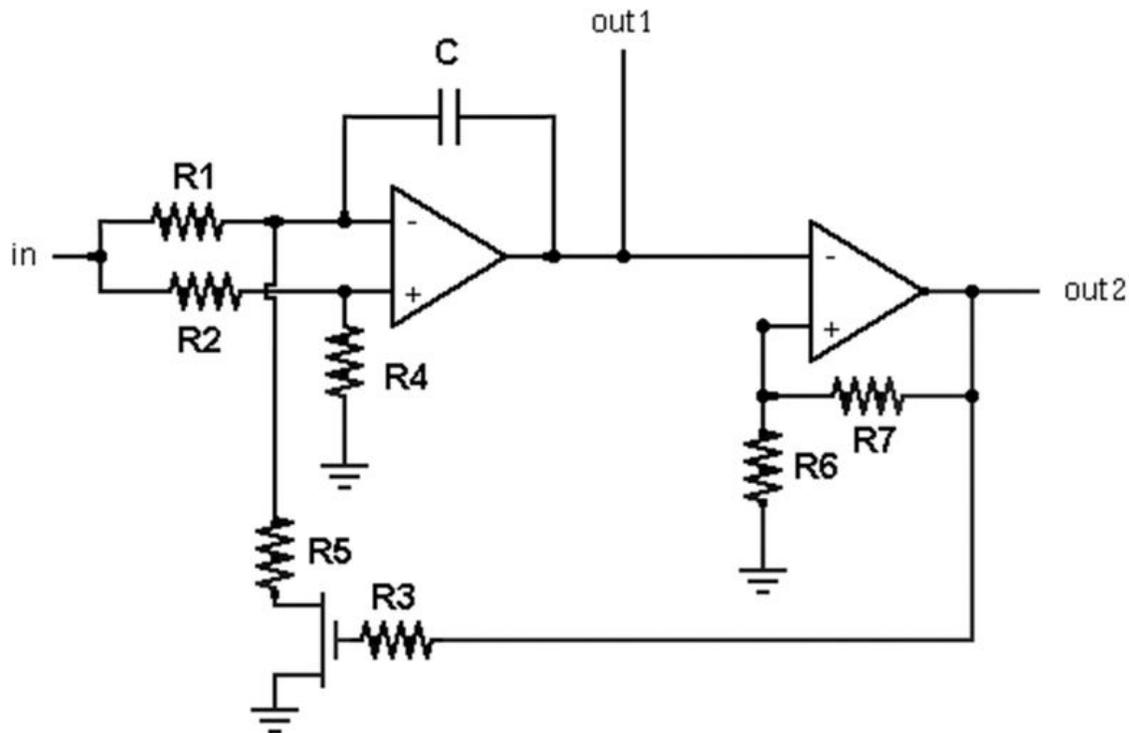
oscillator VCOs can have three topologies: 1) grounded-capacitor VCOs, 2) emitter-coupled VCOs, and 3) delay-based ring VCOs. The first two of these types operate similarly. The amount of time in each state depends on the time for a current to charge or discharge a capacitor. The delay-based ring VCO operates somewhat differently however. For this type, the gain stages are connected in a ring. The output frequency is then a function of the delay in each of stages.

Harmonic oscillator VCOs have these advantages over relaxation oscillators.

- Frequency stability with respect to temperature, noise, and power supply is much better for harmonic oscillator VCOs.
- They have good accuracy for frequency control since the frequency is controlled by a crystal or tank circuit.

A disadvantage of harmonic oscillator VCOs is that they cannot be easily implemented in monolithic ICs. Relaxation oscillator VCOs are better suited for this technology. Relaxation VCOs are also tunable over a wider range of frequencies.

Control of frequency in VCOs



Voltage-controlled oscillator schematic - audio

A voltage-controlled capacitor is one method of making an LC oscillator vary its frequency in response to a control voltage. Any reverse-biased semiconductor diode displays a measure of voltage-dependent capacitance and can be used to change the frequency of an oscillator by varying a control voltage applied to the diode. Special-

purpose variable capacitance varactor diodes are available with well-characterized wide-ranging values of capacitance. Such devices are very convenient in the manufacture of voltage-controlled oscillators. For low-frequency VCOs, other methods of varying the frequency (such as altering the charging rate of a capacitor by means of a voltage controlled current source) are used.

The frequency of a ring oscillator is controlled by varying either the supply voltage or the capacitive loading on each stage.

Voltage-controlled crystal oscillators

A **voltage-controlled crystal oscillator (VCXO)** is used when the frequency of operation needs to be adjusted only finely. The frequency of a voltage-controlled crystal oscillator can be varied only by typically a few tens of parts per million (ppm), because the high Q factor of the crystals allows "pulling" over only a small range of frequencies.

There are two reasons for using a VCXO:

- To adjust the output frequency to match (or perhaps be some exact multiple of) an accurate external reference.
- Where the oscillator drives equipment that may generate radio-frequency interference, adding a varying voltage to its control input can disperse the interference spectrum to make it less objectionable.

A **temperature-compensated VCXO (TCVCXO)** incorporates components that partially correct the dependence on temperature of the resonant frequency of the crystal. A smaller range of voltage control then suffices to stabilize the oscillator frequency in applications where temperature varies, such as heat buildup inside a transmitter.

VCO time-domain equations

$$f_{tuning}(t) = K_o \cdot v_{in}(t)$$

$$\int f_{tuning}(t) dt = \theta_{out}(t)$$

- K_o is called the oscillator gain. Its units are hertz per volt.
- $f_{tuning}(t)$ is the symbol for the time-domain waveform that is the VCO's tunable frequency component.
- $\theta_{out}(t)$ is the symbol for the time-domain waveform that is the VCO's output phase.
- $v_{in}(t)$ is the time-domain symbol of the control (input) voltage of the VCO; it is sometimes also represented as $v_{tune}(t)$

VCO freq-domain equations

$$F_{tuning}(s) = K_o \cdot V_{in}(s)$$

$$\frac{F_{tuning}(s)}{s} = \Theta_{out}(s)$$

VCO design and circuits

Tuning range, tuning gain and phase noise are the most important factors of the basic design of a VCO. Generally low phase noise is preferred in the VCO. The important elements that determine the phase noise of an oscillator are the material, transistor's flicker noise corner frequency, the loaded Q of the resonator and the final signal to noise ratio.

Most commonly used VCO circuits are the Clapp and Colpitts oscillators. The more widely used oscillator of the two is Colpitts and these oscillators are very similar in configuration.

VCOs generally have the lowest Q-factor of the used oscillators, and so suffer more jitter than the other types. The jitter can be made low enough for many applications (such as driving an ASIC), in which case VCOs enjoy the advantages of having no off-chip components (expensive) or on-chip inductors (low yields on generic CMOS processes). These oscillators also have larger tuning ranges than the other kinds, which improves yield and is sometimes a feature of the end product (for instance, the dot clock on a graphics card which drives a wide range of monitors).

Applications

VCOs are used in:

- Electronic jamming equipment
- Function generators,
- The production of electronic music, to generate variable tones,
- Phase-locked loops,
- Frequency synthesizers used in communication equipment.

Voltage-Controlled Crystal Oscillator as a Clock Generator

A clock generator is an oscillator that provides a timing signal to synchronize operations in digital circuits. VCXO clock generators are used in many areas such as digital TV, modems, transmitters and computers. Design parameters for a VCXO clock generator are tuning voltage range, center frequency, frequency tuning range and the timing jitter of the output signal. Jitter is a form of phase noise that must be minimised in applications such as radio receivers, transmitters and measuring equipment.

The tuning range of a VCXO is typically a few ppm (parts per million) corresponding to a control voltage range of typically 0 to 3 volts. When a wider selection of clock frequencies is needed the VCXO output can be passed through digital divider circuits to obtain lower frequency(ies) or be fed to a PLL (Phase Locked Loop). ICs containing both

a VCXO (for external crystal) and a PLL are available. A typical application is to provide clock frequencies in a range from 12 kHz to 96 kHz to an audio digital to analog converter.

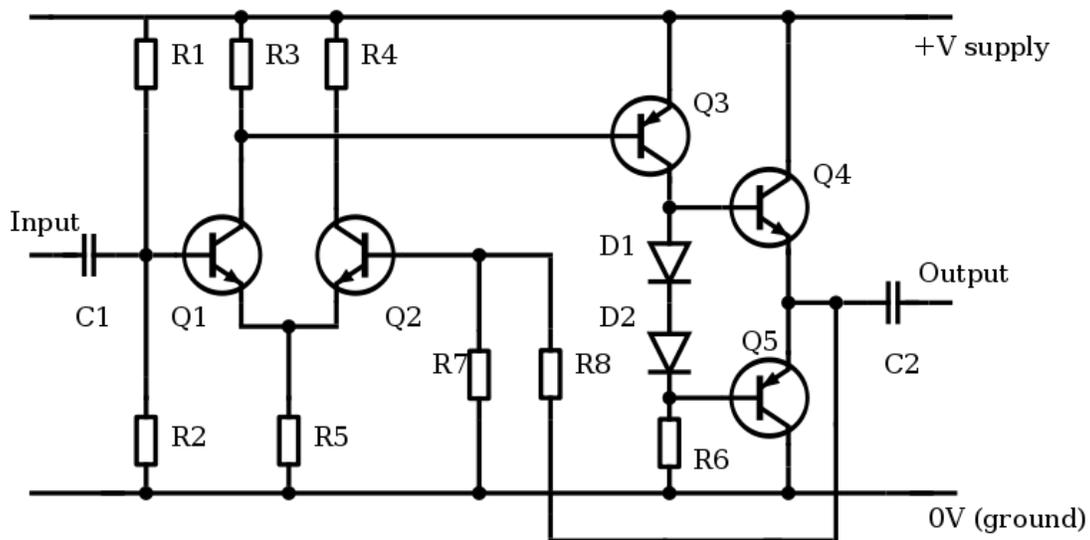
Chips

- LM231, LM331 are VCOs (Precision Voltage-to-Frequency Converters, NS).
- VCOs can be found in the TTL 74124 Dual VCO and the CMOS 4046 PLL.

Chapter 5

Electronic Amplifier

An **electronic amplifier** is a device for increasing the power of a signal. It does this by taking energy from a power supply and controlling the output to match the input signal shape but with a larger amplitude. In this sense, an amplifier may be considered as modulating the output of the power supply.



A practical amplifier circuit

Types of amplifier

Amplifiers can be specified according to their input and output properties. They have some kind of gain, or multiplication factor relating the magnitude of the output signal to the input signal. The gain may be specified as the ratio of output voltage to input voltage

(voltage gain), output power to input power (power gain), or some combination of current, voltage and power. In many cases, with input and output in the same units, gain will be unitless (although often expressed in decibels); for others this is not necessarily so. For example, a transconductance amplifier has a gain with units of conductance (output current per input voltage). The power gain of an amplifier depends on the source and load impedances used as well as its voltage gain; while an RF amplifier may have its impedances optimized for power transfer, audio and instrumentation amplifiers are normally employed with amplifier input and output impedances optimized for least loading and highest quality. So an amplifier that is said to have a gain of 20 dB might have a voltage gain of ten times and an available power gain of much more than 20 dB (100 times power ratio), yet be delivering a much lower power gain if, for example, the input is a 600 ohm microphone and the output is a 47 kilohm power amplifier's input socket.

In most cases an amplifier should be linear; that is, the gain should be constant for any combination of input and output signal. If the gain is not constant, e.g., by clipping the output signal at the limits of its capabilities, the output signal will be distorted. There are however cases where variable gain is useful.

There are many types of electronic amplifiers, commonly used in radio and television transmitters and receivers, high-fidelity ("hi-fi") stereo equipment, microcomputers and other electronic digital equipment, and guitar and other instrument amplifiers. Critical components include active devices, such as vacuum tubes or transistors. A brief introduction to the many types of electronic amplifier follows.

Power amplifier

The term "power amplifier" is a relative term with respect to the amount of power delivered to the load and/or sourced by the supply circuit. In general a power amplifier is designated as the last amplifier in a transmission chain (the *output stage*) and is the amplifier stage that typically requires most attention to power efficiency. Efficiency considerations lead to various classes of power amplifier.

Vacuum tube (valve) amplifiers



The glow from four "Electro Harmonix KT88" brand power tubes lights up the inside of a Traynor YBA-200 guitar amplifier

According to Symons, while semiconductor amplifiers have largely displaced valve amplifiers for low power applications, valve amplifiers are much more cost effective in high power applications such as "radar, countermeasures equipment, or communications equipment" (p. 56). Many microwave amplifiers are specially designed valves, such as the klystron, gyrotron, traveling wave tube, and crossed-field amplifier, and these microwave valves provide much greater single-device power output at microwave frequencies than solid-state devices (p. 59).

Valves/tube amplifiers also have niche uses in other areas, such as

- Electric guitar amplification
- in Russian military aircraft, for their EMP tolerance
- niche audio for their sound qualities (recording, and audiophile equipment)

Transistor amplifiers

The essential role of this active element is to magnify an input signal to yield a significantly larger output signal. The amount of magnification (the "forward gain") is determined by the external circuit design as well as the active device.

Many common active devices in transistor amplifiers are bipolar junction transistors (BJTs) and metal oxide semiconductor field-effect transistors (MOSFETs).

Applications are numerous, some common examples are audio amplifiers in a home stereo or PA system, RF high power generation for semiconductor equipment, to RF and Microwave applications such as radio transmitters.

Transistor-based amplifier can be realized using various configurations: for example with a bipolar junction transistor we can realize common base, common collector or common emitter amplifier; using a MOSFET we can realize common gate, common source or common drain amplifier. Each configuration has different characteristic (gain, impedance...).

Operational amplifiers (op-amps)

An operational amplifier is an amplifier circuit with very high open loop gain and differential inputs which employs external feedback for control of its transfer function or gain. Although the term is today commonly applied to integrated circuits, the original operational amplifier design was implemented with valves.

Fully differential amplifiers (FDA)

A fully differential amplifier is a solid state integrated circuit amplifier which employs external feedback for control of its transfer function or gain. It is similar to the operational amplifier but it also has differential output pins.

Video amplifiers

These deal with video signals and have varying bandwidths depending on whether the video signal is for SDTV, EDTV, HDTV 720p or 1080i/p etc.. The specification of the bandwidth itself depends on what kind of filter is used and which point (-1 dB or -3 dB for example) the bandwidth is measured. Certain requirements for step response and overshoot are necessary in order for acceptable TV images to be presented.

Oscilloscope vertical amplifiers

These are used to deal with video signals to drive an oscilloscope display tube and can have bandwidths of about 500 MHz. The specifications on step response, rise time, overshoot and aberrations can make the design of these amplifiers extremely difficult. One of the pioneers in high bandwidth vertical amplifiers was the Tektronix company.

Distributed amplifiers

These use transmission lines to temporally split the signal and amplify each portion separately in order to achieve higher bandwidth than can be obtained from a single amplifying device. The outputs of each stage are combined in the output transmission line. This type of amplifier was commonly used on oscilloscopes as the final vertical amplifier. The transmission lines were often housed inside the display tube glass envelope.

Switched mode amplifiers

These nonlinear amplifiers have much higher efficiencies than linear amps, and are used where the power saving justifies the extra complexity.

Negative resistance devices

Negative resistances can be used as amplifiers, such as the tunnel diode amplifier.

Microwave amplifiers

Travelling wave tube amplifiers

Traveling wave tube amplifiers (TWTAs) are used for high power amplification at low microwave frequencies. They typically can amplify across a broad spectrum of frequencies; however, they are usually not as tunable as klystrons.

Klystrons

Klystrons are vacuum-devices that do not have as wide a bandwidth as TWTAs. They generally are also much heavier than TWTAs, and are therefore ill-suited for light-weight mobile applications. Klystrons are tunable, offering selective output within their specified frequency range.

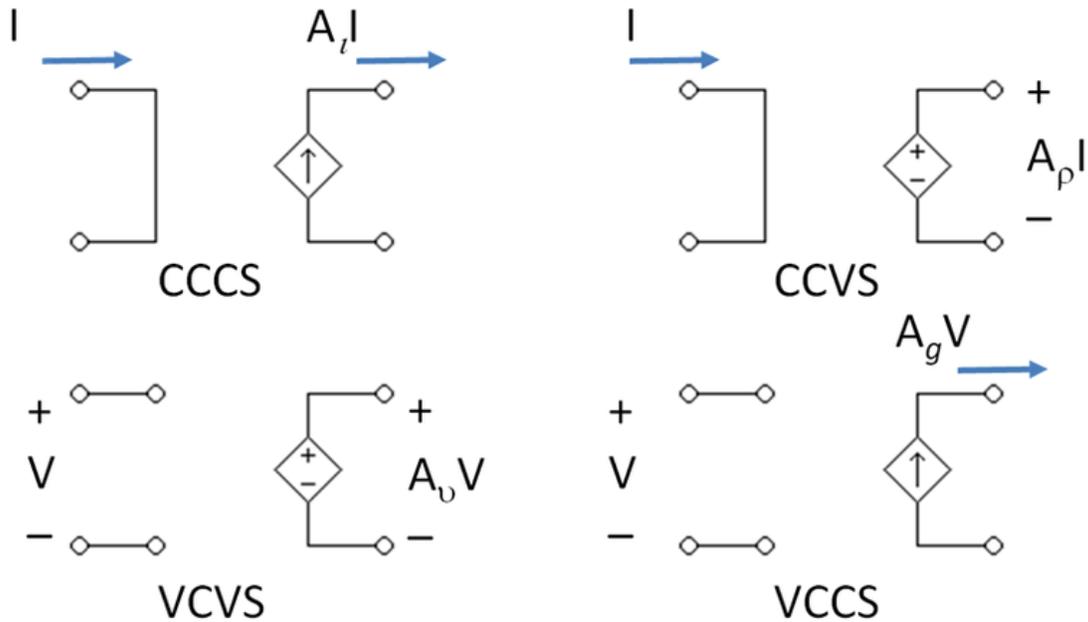
Musical instrument (audio) amplifiers

An audio amplifier is usually used to amplify signals such as music or speech.

Classification of amplifier stages and systems

There are many alternative classifications that address different aspects of amplifier designs, and they all express some particular perspective relating the design parameters to the objectives of the circuit. Amplifier design is always a compromise of numerous factors, such as cost, power consumption, real-world device imperfections, and a multitude of performance specifications. Below are several different approaches to classification:

Input and output variables



The four types of dependent source; control variable on left, output variable on right

Electronic amplifiers use two variables: current and voltage. Either can be used as input, and either as output leading to four types of amplifiers. In idealized form they are represented by each of the four types of dependent source used in linear analysis, as shown in the figure, namely:

Input	Output	Dependent source	Amplifier type
I	I	current controlled current source CCCS	current amplifier
I	V	current controlled voltage source CCVS	transresistance amplifier
V	I	voltage controlled current source VCCS	transconductance amplifier
V	V	voltage controlled voltage source VCVS	voltage amplifier

Each type of amplifier in its ideal form has an ideal input and output resistance that is the same as that of the corresponding dependent source:

Amplifier type	Dependent source	Input impedance	Output impedance
Current	CCCS	0	∞
Transresistance	CCVS	0	0
Transconductance	VCCS	∞	∞
Voltage	VCVS	∞	0

In practice the ideal impedances are only approximated. For any particular circuit, a small-signal analysis is often used to find the impedance actually achieved. A small-

signal AC test current I_x is applied to the input or output node, all external sources are set to AC zero, and the corresponding alternating voltage V_x across the test current source determines the impedance seen at that node as $R = V_x / I_x$.

Amplifiers designed to attach to a transmission line at input and/or output, especially RF amplifiers, do not fit into this classification approach. Rather than dealing with voltage or current individually, they ideally couple with an input and/or output impedance matched to the transmission line impedance, that is, match *ratios* of voltage to current. Many real RF amplifiers come close to this ideal. Although, for a given appropriate source and load impedance, RF amplifiers can be characterized as amplifying voltage or current, they fundamentally are amplifying power.

Common terminal

One set of classifications for amplifiers is based on which device terminal is common to both the input and the output circuit. In the case of bipolar junction transistors, the three classes are common emitter, common base, and common collector. For field-effect transistors, the corresponding configurations are common source, common gate, and common drain; for triode vacuum devices, *common cathode*, *common grid*, and *common plate*. The output voltage of a common plate amplifier is the same as the input (this arrangement is used as the input presents a high impedance and does not load the signal source, although it does not amplify the voltage), i.e., the output at the cathode follows the input at the grid; consequently it was commonly called a *cathode follower*. By analogy the terms *emitter follower* and *source follower* are sometimes used.

Unilateral or bilateral

When an amplifier has an output that exhibits no feedback to its input side, it is called **unilateral**. The input impedance of a unilateral amplifier is independent of the load, and the output impedance is independent of the signal source impedance.

If feedback connects part of the output back to the input of the amplifier it is called a **bilateral** amplifier. The input impedance of a bilateral amplifier is dependent upon the load, and the output impedance is dependent upon the signal source impedance.

All amplifiers are bilateral to some degree; however they may often be modeled as unilateral under operating conditions where feedback is small enough to neglect for most purposes, simplifying analysis.

Negative feedback is often applied deliberately to tailor amplifier behavior. Some feedback, which may be positive or negative, is unavoidable and often undesirable, introduced, for example, by parasitic elements such as the inherent capacitance between input and output of a device such as a transistor and capacitive coupling due to external wiring. Excessive frequency-dependent positive feedback may cause what is supposed to be an amplifier to become an oscillator.

Linear unilateral and bilateral amplifiers can be represented as two-port networks.

Inverting or non-inverting

Another way to classify amps is the phase relationship of the input signal to the output signal. An **inverting** amplifier produces an output 180 degrees out of phase with the input signal (that is, a polarity inversion or mirror image of the input as seen on an oscilloscope). A **non-inverting** amplifier maintains the phase of the input signal waveforms. An **emitter follower** is a type of non-inverting amplifier, indicating that the signal at the emitter of a transistor is following (that is, matching with unity gain but perhaps an offset) the input signal.

This description can apply to a single stage of an amplifier, or to a complete amplifier system.

Function

Other amplifiers may be classified by their function or output characteristics. These functional descriptions usually apply to complete amplifier systems or sub-systems and rarely to individual stages.

- A **servo amplifier** indicates an integrated feedback loop to actively control the output at some desired level. A **DC servo** indicates use at frequencies down to DC levels, where the rapid fluctuations of an audio or RF signal do not occur. These are often used in mechanical actuators, or devices such as DC motors that must maintain a constant speed or torque. An **AC servo** amp can do this for some ac motors.
- A **linear** amplifier responds to different frequency components independently, and does not generate harmonic distortion or Intermodulation distortion (well, hardly any). A **nonlinear** amplifier does generate distortion (e.g. the output is a current to a lamp that must be either fully on or off, but the input is continuously variable; or the amplifier is used in an analog computer where a special transfer function, such as logarithmic, is desired; or a following tuned circuit will remove the harmonics generated by a non-linear RF amplifier).
- A **wideband** amplifier has a precise amplification factor over a wide range of frequencies, and is often used to boost signals for relay in communications systems. A **narrowband** amp is made to amplify only a specific narrow range of frequencies, to the exclusion of other frequencies.
- An **RF** amplifier refers to an amplifier designed for use in the radio frequency range of the electromagnetic spectrum, and is often used to increase the sensitivity of a receiver or the output power of a transmitter.
- An **audio amplifier** is designed for use in reproducing audio frequencies. This category subdivides into small signal amplification, and power amps which are optimised for driving speakers, sometimes with multiple amps grouped together as separate or bridgeable channels to accommodate different audio reproduction requirements. Frequently used terms within audio amplifiers include:

- preamplifier (preamp), that may include phono or gramophone preamp with equalization for RIAA LP recordings, or tape head preamps with CCIR equalisation filters; they may include filters or tone control circuitry.
- power amplifier (normally assumed to drive loudspeakers), headphone amplifiers, and public address amplifiers.
- stereo amplifiers imply two channels of output (left and right), although the term simply means "solid" sound (referring to three-dimensional) - so quadraphonic stereo was used for amplifiers with 4 channels; 5.1 and 7.1 systems refer to Home theatre systems with 5 or 7 normal spacial channels, plus a subwoofer channel (that is not very directional).
- Buffer amplifiers, that may include emitter followers, provide a high impedance input for a device (perhaps another amplifier, or perhaps an energy-hungry load such as lights) that would otherwise draw too much current from the source. Line drivers are a type of buffer intended to feed long or interference-prone interconnect cables, possibly with differential outputs if driving twisted pairs of cables.
- A special type of amplifier is widely used in instruments and for signal processing, among many other varied uses. These are known as **operational amplifiers** or **op-amps**. This is because this type of amplifier is used in circuits that perform mathematical algorithmic functions, or "operations" on input signals to obtain specific types of output signals. A typical modern op-amp has differential inputs (one "inverting", one "non-inverting") and one output. An idealised op-amp has the following characteristics:
 - Infinite input impedance (so as to not load circuitry it is sampling as a control input)
 - Zero output impedance
 - Infinite gain
 - Zero propagation delay

The performance of an op-amp with these characteristics would be entirely defined by the (usually passive) components forming a negative feedback loop around it, that is, *the amplifier itself has no effect on the output*.

Today, op-amps are usually provided as integrated circuits, rather than constructed from discrete components. All real-world op-amps fall short of the idealised specification above – but some modern components have remarkable performance and come close in some respects.

Interstage coupling method

Amplifiers are sometimes classified by the coupling method of the signal at the input, output, or between stages. Different types of these include:

Resistive-capacitive (RC) coupled amplifier, using a network of resistors and capacitors
 By design these amplifiers cannot amplify DC signals as the capacitors block the DC component of the input signal. RC-coupled amplifiers were used very often in

circuits with vacuum tubes or discrete transistors. In the days of the integrated circuit a few more transistors on a chip are much cheaper and smaller than a capacitor.

Inductive-capacitive (LC) coupled amplifier, using a network of inductors and capacitors

This kind of amplifier is most often used in selective radio-frequency circuits.

Transformer coupled amplifier, using a transformer to match impedances or to decouple parts of the circuits

Quite often LC-coupled and transformer-coupled amplifiers cannot be distinguished as a transformer is some kind of inductor.

Direct coupled amplifier, using no impedance and bias matching components

This class of amplifier was very uncommon in the vacuum tube days when the anode (output) voltage was at greater than several hundred volts and the grid (input) voltage at a few volts minus. So they were only used if the gain was specified down to DC (e.g., in an oscilloscope). In the context of modern electronics developers are encouraged to use directly coupled amplifiers whenever possible.

Frequency range

Depending on the frequency range and other properties amplifiers are designed according to different principles.

- Frequency ranges down to DC are only used when this property is needed. DC amplification leads to specific complications that are avoided if possible; **DC-blocking** capacitors are added to remove DC and sub-sonic frequencies from audio amplifiers.
- Depending on the frequency range specified different design principles must be used. Up to the MHz range only "discrete" properties need be considered; e.g., a terminal has an input impedance.
- As soon as any connection within the circuit gets longer than perhaps 1% of the wavelength of the highest specified frequency (e.g., at 100 MHz the wavelength is 3 m, so the critical connection length is approx. 3 cm) design properties radically change. For example, a specified length and width of a PCB trace can be used as a selective or impedance-matching entity.
- Above a few 100 MHz, it gets difficult to use discrete elements, especially inductors. In most cases PCB traces of very closely defined shapes are used instead.

The frequency range handled by an amplifier might be specified in terms of bandwidth (normally implying a response that is 3 dB down when the frequency reaches the specified bandwidth), or by specifying a frequency response that is within a certain number of decibels between a lower and an upper frequency (e.g. "20 Hz to 20 kHz plus or minus 1 dB").

Type of load

- Untuned
 - audio
 - video
- Tuned (RF amps) - used for amplifying a single radio frequency or a band of frequencies

Implementation

Amplifiers are implemented using active elements of different kinds:

- The first active elements were relays. They were for example used in transcontinental telegraph lines: a weak current was used to switch the voltage of a battery to the outgoing line.
- For transmitting audio, carbon microphones were used as the active element. This was used to modulate a radio-frequency source in one of the first AM audio transmissions, by Reginald Fessenden on Dec. 24, 1906.
- In the 1960s, the transistor started to take over. These days, discrete transistors are still used in high-power amplifiers and in specialist audio devices.
- Up to the early 1970s, most amplifiers used vacuum tubes ("valves" in the UK). Today, tubes are used for specialist audio applications such as guitar amplifiers and audiophile amplifiers. Many broadcast transmitters still use vacuum tubes.
- Beginning in the 1970s, more and more transistors were connected on a single chip therefore creating the integrated circuit. Nearly all amplifiers commercially available today are based on integrated circuits.

For exotic purposes, other active elements have been used. For example, in the early days of the communication satellite parametric amplifiers were used. The core circuit was a diode whose capacity was changed by an RF signal created locally. Under certain conditions, this RF signal provided energy that was modulated by the extremely weak satellite signal received at the earth station. The operating principle of a parametric amplifier is somewhat similar to the principle by which children keep their swings in motion: as long as the swing moves you only need to change a parameter of the swinging entity; e.g., you must move your center of gravity up and down. In our case, the capacity of the diode is changed periodically.

Power amplifier classes

Angle of flow or conduction angle

Power amplifier circuits (output stages) are classified as A, B, AB and C for analog designs, and class D and E for switching designs based upon the conduction angle or *angle of flow*, Θ , of the input signal through the (or each) output amplifying device, that is, the portion of the input signal cycle during which the amplifying device conducts. The image of the conduction angle is derived from amplifying a sinusoidal signal. (If the

device is always on, $\Theta = 360^\circ$.) The angle of flow is closely related to the amplifier power efficiency. The various classes are introduced below, followed by more detailed discussion under individual headings later on.

Class A

100% of the input signal is used (conduction angle $\Theta = 360^\circ$ or 2π); i.e., the active element remains conducting (works in its "linear" range) all of the time. Where efficiency is not a consideration, most small signal linear amplifiers are designed as class A. Class A amplifiers are typically more linear and less complex than other types, but are very inefficient. This type of amplifier is most commonly used in small-signal stages or for low-power applications (such as driving headphones). Subclass A2 is sometimes used to refer to vacuum tube class A stages where the grid is allowed to be driven slightly positive on signal peaks, resulting in slightly more power than normal class A (A1; where the grid is always negative), but incurring more distortion.

Class B

50% of the input signal is used ($\Theta = 180^\circ$ or π ; i.e., the active element works in its linear range half of the time and is more or less turned off for the other half). In most class B, there are two output devices (or sets of output devices), each of which conducts alternately (push-pull) for exactly 180° (or half cycle) of the input signal; selective RF amplifiers can also be implemented using a single active element.

These amplifiers are subject to *crossover distortion* if the transition from one active element to the other is not perfect, as when two complementary transistors (i.e., one PNP, one NPN) are connected as two emitter followers with their base and emitter terminals in common, requiring the base voltage to slew across the region where both devices are turned off.

Class AB

Here the two active elements conduct more than half of the time as a means to reduce the cross-over distortions of class B amplifiers. In the example of the complementary emitter followers a bias network allows for more or less quiescent current thus providing an operating point somewhere between class A and class B. Sometimes a figure is added (e.g., AB₁ or AB₂) for vacuum tube stages where the grid voltage is always negative with respect to the cathode (class AB₁) or may be slightly positive (hence drawing grid current, adding more distortion, but giving slightly higher output power) on signal peaks (class AB₂); another interpretation being higher figures implying a higher quiescent current and therefore more of the properties of class A.

Class C

Less than 50% of the input signal is used (conduction angle $\Theta < 180^\circ$). The advantage is potentially high efficiency, but a disadvantage is high distortion.

Class D

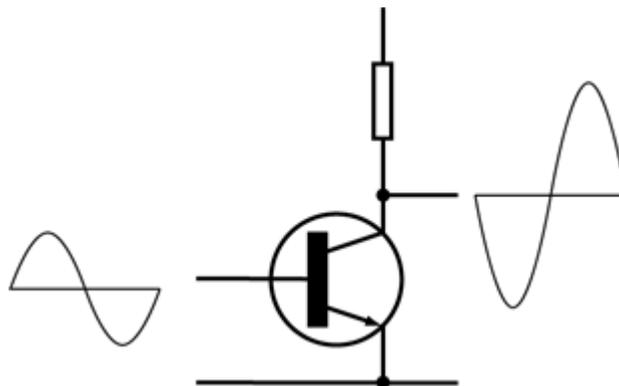
These use switching to achieve a very high power efficiency (more than 90% in modern designs). By allowing each output device to be either fully on or off, losses are minimized. The analog output is created by pulse-width modulation; i.e., the active element is switched on for shorter or longer intervals instead of modifying its resistance. There are more complicated switching schemes like sigma-delta modulation, to improve some performance aspects like lower distortions or better efficiency.

Additional classes

There are several other amplifier classes, although they are mainly variations of the previous classes. For example, class G and class H amplifiers are marked by variation of the supply rails (in discrete steps or in a continuous fashion, respectively) following the input signal. Wasted heat on the output devices can be reduced as excess voltage is kept to a minimum. The amplifier that is fed with these rails itself can be of any class. These kinds of amplifiers are more complex, and are mainly used for specialized applications, such as very high-power units. Also, class E and class F amplifiers are commonly described in literature for radio frequencies applications where efficiency of the traditional classes in are important, yet several aspects not covered elsewhere (e.g.: amplifiers often simply said to have a gain of x dB - so what power gain?) deviate substantially from their ideal values. These classes use harmonic tuning of their output networks to achieve higher efficiency and can be considered a subset of Class C due to their conduction angle characteristics.

The classes can be most easily understood using the diagrams in each section below. For the sake of illustration, a bipolar junction transistor is shown as the amplifying device, but in practice this could be a MOSFET or vacuum tube device. In an analog amplifier (the most common kind), the signal is applied to the input terminal of the device (base, gate or grid), and this causes a proportional output drive current to flow out of the output terminal. The output drive current comes from the power supply.

Class A



Class A amplifier

Amplifying devices operating in class A conduct over the whole of the input cycle such that the output signal is an exact scaled-up replica of the input with no clipping. A *class A amplifier* is distinguished by the *output stage* being biased into class A.

Advantages of class A amplifiers

- Class A designs are simpler than other classes; for example class AB and B designs require two devices (push-pull output) to handle both halves of the waveform; class A can use a single device single-ended.
- The amplifying element is biased so the device is always conducting to some extent, normally implying the quiescent (small-signal) collector current (for transistors; drain current for FETs or anode/plate current for vacuum tubes) is close to the most linear portion of its transconductance curve.
- Because the device is never shut off completely there is no "turn on" time, little problem with charge storage, and generally better high frequency performance and feedback loop stability (and usually fewer high-order harmonics).
- The point at which the device comes closest to being cut off is not close to zero signal, so the problem of crossover distortion associated with class AB and B designs is avoided.

Disadvantage of class A amplifiers

- They are very inefficient; a theoretical maximum of 50% is obtainable with inductive output coupling and only 25% with capacitive coupling, unless Square law output stages are used. In a power amplifier this not only wastes power and limits battery operation, it may place restrictions on the output devices that can be used (for example: ruling out some audio triodes if modern low-efficiency loudspeakers are to be used), and will increase costs. Inefficiency comes not just from the fact that the device is always conducting to some extent (that happens even with class AB, yet its efficiency can be close to that of class B); it is that the standing current is roughly half the maximum output current (although this can be less with Square law output stage), together with the problem that a large part of the power supply voltage is developed across the output device at low signal levels (as with classes AB and B, but unlike output stages such as class D). If high output powers are needed from a class A circuit, the power waste (and the accompanying heat) will become significant. For every watt delivered to the load, the amplifier itself will, *at best*, dissipate another watt. For large powers this means very large and expensive power supplies and heat sinking.

Class A designs have largely been superseded by the more efficient designs for power amplifiers, though they remain popular with some hobbyists, mostly for their simplicity. Also, many audiophiles believe that class A gives the best sound quality (for their absence of crossover distortion and reduced odd-harmonic and high-order harmonic distortion) which provides a small market for expensive **high fidelity** class A amps.

Single-ended and triode class A amplifiers

Some aficionados who prefer class A amplifiers also prefer the use of thermionic valve (or "tube") designs instead of transistors, especially in Single-ended triode output configurations for several claimed reasons:

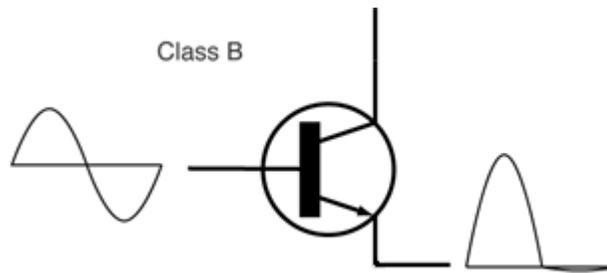
- Single-ended output stages (be they tube or transistor) have an asymmetrical transfer function, meaning that even harmonics in the created distortion tend not to be canceled (as they are in push-pull output stages); by using tubes OR FETs most of the distortion is from the square law transfer characteristic and so second-order, which some consider to be "warmer" and more pleasant.
- For those who prefer low distortion figures, the use of tubes with class A (generating little odd-harmonic distortion, as mentioned above) together with symmetrical circuits (such as push-pull output stages, or balanced low-level stages) results in the cancellation of most of the even distortion harmonics, hence the removal of most of the distortion.
- Though good amplifier design can reduce harmonic distortion patterns to almost nothing, distortion is essential to the sound of electric guitar amplifiers, for example, and is held by recording engineers to offer more flattering microphones and to enhance "clinical-sounding" digital technology.
- Historically, valve amplifiers often used a class A power amplifier simply because valves are large and expensive; many class A designs use only a single device.

Transistors are much cheaper, and so more elaborate designs that give greater efficiency but use more parts are still cost-effective. A classic application for a pair of class A devices is the long-tailed pair, which is exceptionally linear, and forms the basis of many more complex circuits, including many audio amplifiers and almost all op-amps. Class A amplifiers are often used in output stages of high quality op-amps (although the accuracy of the bias in low cost op-amps such as the **741** may result in class A or class AB or class B, varying from device to device or with temperature). They are sometimes used as medium-power, low-efficiency, and high-cost audio amplifiers. The power consumption is unrelated to the output power. At idle (no input), the power consumption is essentially the same as at high output volume. The result is low efficiency and high heat dissipation.

Class B and AB

Class B or AB push-pull circuits are the most common design type found in audio power amplifiers. Class AB is widely considered a good compromise for audio amplifiers, since much of the time the music is quiet enough that the signal stays in the "class A" region, where it is amplified with good fidelity, and by definition if passing out of this region, is large enough that the distortion products typical of class B are relatively small. The crossover distortion can be reduced further by using negative feedback. Class B and AB amplifiers are sometimes used for RF linear amplifiers as well. Class B amplifiers are also favored in battery-operated devices, such as transistor radios.

Class B



Class B amplifier

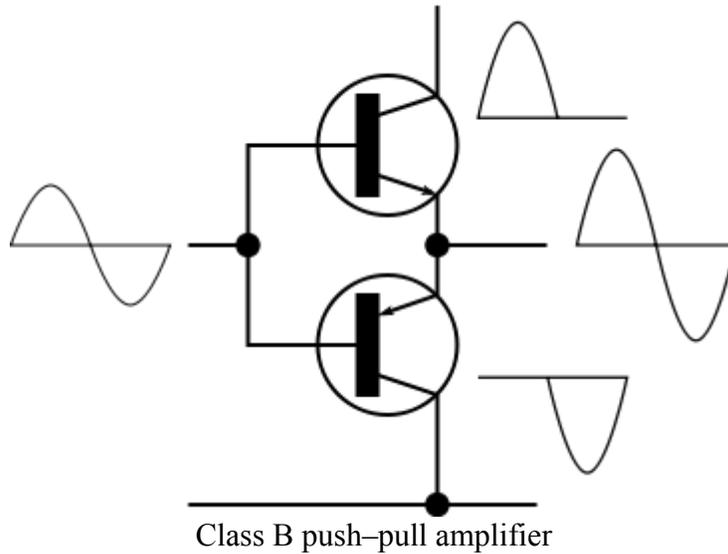
Class B amplifiers only amplify half of the input wave cycle, thus creating a large amount of distortion, but their efficiency is greatly improved and is much better than class A. Class B has a maximum theoretical efficiency of 78.5% (i.e., $\pi/4$). This is because the amplifying element is switched off altogether half of the time, and so cannot dissipate power. A single class B element is rarely found in practice, though it has been used for driving the loudspeaker in the early IBM Personal Computers with beeps, and it can be used in RF power amplifier where the distortion levels are less important. However, class C is more commonly used for this.

A practical circuit using class B elements is the push-pull stage, such as the very simplified complementary pair arrangement shown below. Here, complementary or quasi-complementary devices are each used for amplifying the opposite halves of the input signal, which is then recombined at the output. This arrangement gives excellent efficiency, but can suffer from the drawback that there is a small mismatch in the crossover region - at the "joins" between the two halves of the signal, as one output device has to take over supplying power exactly as the other finishes. This is called crossover distortion. An improvement is to bias the devices so they are not completely off when they're not in use. This approach is called *class AB* operation.

Digital class B

A limited power output class B amplifier with a single-ended supply rail of 5 ± 0.5 V.

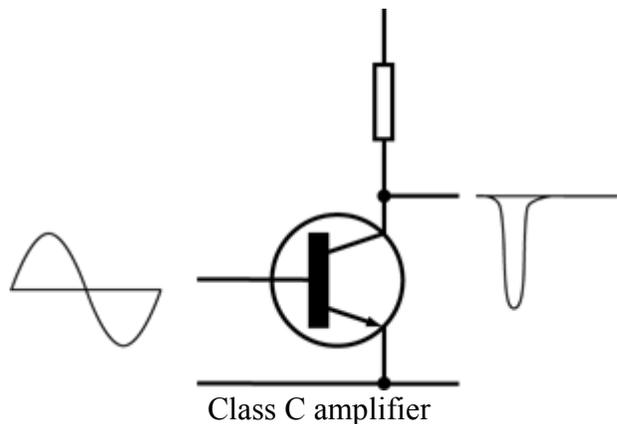
Class AB



In class AB operation, each device operates the same way as in class B over half the waveform, but also conducts a small amount on the other half. As a result, the region where both devices simultaneously are nearly off (the "dead zone") is reduced. The result is that when the waveforms from the two devices are combined, the crossover is greatly minimised or eliminated altogether. The exact choice of **quiescent current**, the standing current through both devices when there is no signal, makes a large difference to the level of distortion (and to the risk of thermal runaway, that may damage the devices); often the bias voltage applied to set this quiescent current has to be adjusted with the temperature of the output transistors. Another approach (often used as well as thermally-tracking bias voltages) is to include small value resistors in series with the emitters.

Class AB sacrifices some efficiency over class B in favor of linearity, thus is less efficient (below 78.5% for full-amplitude sinewaves in transistor amplifiers, typically; much less is common in class AB vacuum tube amplifiers). It is typically much more efficient than class A.

Class C

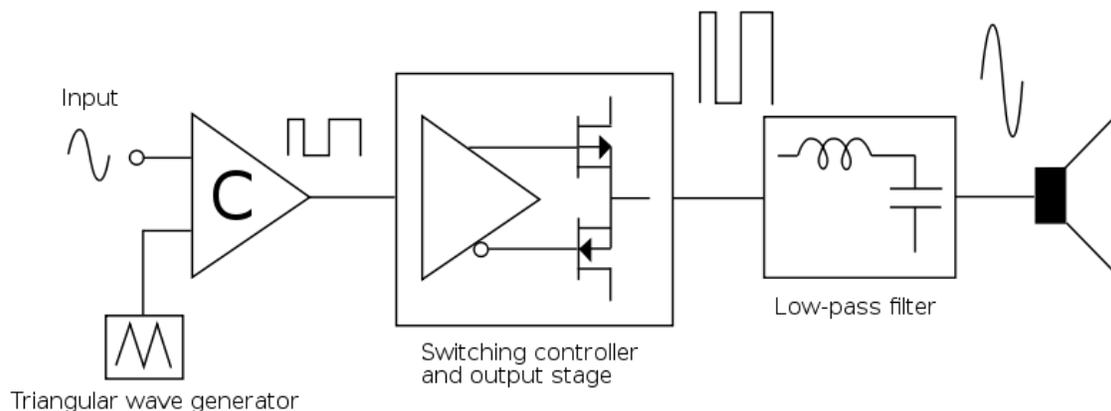


Class C amplifiers conduct less than 50% of the input signal and the distortion at the output is high, but high efficiencies (up to 90%) are possible. Some applications (for example, megaphones) can tolerate the distortion. A much more common application for class C amplifiers is in RF transmitters, where the distortion can be vastly reduced by using tuned loads on the amplifier stage. The input signal is used to roughly switch the amplifying device on and off, which causes pulses of current to flow through a tuned circuit.

The class C amplifier has two modes of operation: tuned and untuned. The diagram shows a waveform from a simple class C circuit without the tuned load. This is called untuned operation, and the analysis of the waveforms shows the massive distortion that appears in the signal. When the proper load (e.g., a pure inductive-capacitive filter) is used, two things happen. The first is that the output's bias level is clamped, so that the output variation is centered at one-half of the supply voltage. This is why tuned operation is sometimes called a *clammer*. This action of elevating bias level allows the waveform to be restored to its proper shape, allowing a complete waveform to be re-established despite having only a one-polarity supply. This is directly related to the second phenomenon: the waveform on the center frequency becomes much less distorted. The distortion that is present is dependent upon the bandwidth of the tuned load, with the center frequency seeing very little distortion, but greater attenuation the farther from the tuned frequency that the signal gets.

The tuned circuit will only resonate at particular frequencies, and so the unwanted frequencies are dramatically suppressed, and the wanted full signal (sine wave) will be extracted by the tuned load (e.g., a high-quality bell will ring at a particular frequency when it is hit periodically with a hammer). Provided the transmitter is not required to operate over a very wide band of frequencies, this arrangement works extremely well. Other residual harmonics can be removed using a filter.

Class D



Block diagram of a basic switching or PWM (class D) amplifier.



Boss Audio class D mono car audio amplifier with a low pass filter for powering subwoofers

In the class D amplifier the input signal is converted to a sequence of higher voltage output pulses. The averaged-over-time power values of these pulses are directly proportional to the instantaneous amplitude of the input signal. The frequency of the output pulses is typically ten or more times the highest frequency in the input signal to be amplified. The output pulses contain inaccurate spectral components (that is, the pulse frequency and its harmonics) which must be removed by a low-pass passive filter. The resulting filtered signal is then an amplified replica of the input.

These amplifiers use pulse width modulation, pulse density modulation (sometimes referred to as pulse frequency modulation) or more advanced form of modulation such as Delta-sigma modulation (for example, in the Analog Devices AD1990 class D audio power amplifier). Output stages such as those used in pulse generators are examples of class D amplifiers. The term *class D* is usually applied to devices intended to reproduce signals with a bandwidth well below the switching frequency.

Class D amplifiers can be controlled by either analog or digital circuits. The digital control introduces additional distortion called *quantization error* caused by its conversion of the input signal to a digital value.

The main advantage of a class D amplifier is power efficiency. Because the output pulses have a fixed amplitude, the switching elements (usually MOSFETs, but valves and bipolar transistors were once used) are switched either completely on or completely off, rather than operated in linear mode. A MOSFET operates with the lowest resistance when fully-on and thus has the lowest power dissipation when in that condition, except when fully off. When operated in a linear mode the MOSFET has variable amounts of resistance that vary linearly with the input voltage and the resistance is something other than the minimum possible, therefore more electrical energy is dissipated as heat. Compared to class A/B operation, class D's lower losses permit the use of a smaller heat sink for the MOSFETS while also reducing the amount of AC power supply power required. Thus, class D amplifiers do not need as large or as heavy power supply transformers or heatsinks, so they are smaller and more compact in size than an equivalent class AB amplifier.

Class D amplifiers have been widely used to control motors, and almost exclusively for small DC motors, but they are now also used as audio amplifiers, with some extra circuitry to allow analogue to be converted to a much higher frequency pulse width modulated signal. The relative difficulty of achieving good audio quality means that nearly all are used in applications where quality is not a factor, such as modestly-priced bookshelf audio systems and "DVD-receivers" in mid-price home theater systems.

High quality class D audio amplifiers have now appeared in the market and these revised designs have been said to rival good traditional AB amplifiers in terms of quality. Before these higher quality designs existed an earlier use of class D amplifiers and prolific area of application was high-powered, subwoofer amplifiers in cars. Because subwoofers are generally limited to a bandwidth of no higher than 150 Hz, the switching speed for the amplifier does not have to be as high as for a full range amplifier. Class D amplifiers for driving subwoofers are relatively inexpensive, in comparison to class AB amplifiers.

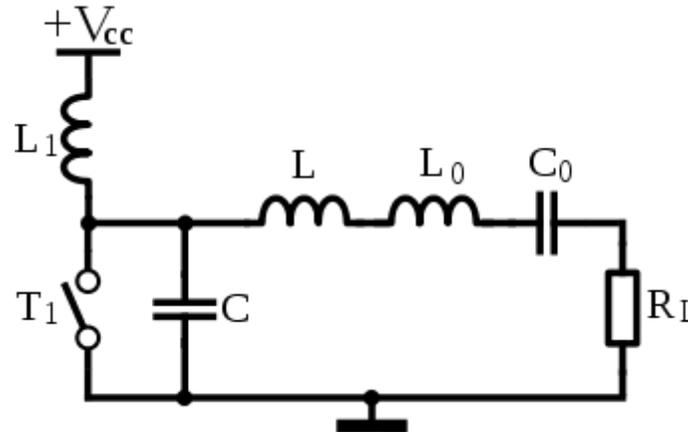
The letter *D* used to designate this amplifier class is simply the next letter after *C*, and does not stand for *digital*. Class D and class E amplifiers are sometimes mistakenly described as "digital" because the output waveform superficially resembles a pulse-train of digital symbols, but a class D amplifier merely converts an input waveform into a continuously pulse-width modulated (square wave) analog signal. (A digital waveform would be pulse-code modulated.)

Additional classes

Class E

The class E/F amplifier is a highly efficient switching power amplifier, typically used at such high frequencies that the switching time becomes comparable to the duty time. As said in the class D amplifier, the transistor is connected via a series LC circuit to the load, and connected via a large L (inductor) to the supply voltage. The supply voltage is connected to ground via a large capacitor to prevent any RF signals leaking into the

supply. The class E amplifier adds a C (capacitor) between the transistor and ground and uses a defined L_1 to connect to the supply voltage.



Class E amplifier

The following description ignores DC, which can be added easily afterwards. The above mentioned C and L are in effect a parallel LC circuit to ground. When the transistor is on, it pushes through the serial LC circuit into the load and some current begins to flow to the parallel LC circuit to ground. Then the serial LC circuit swings back and compensates the current into the parallel LC circuit. At this point the current through the transistor is zero and it is switched off. Both LC circuits are now filled with energy in C and L_0 . The whole circuit performs a damped oscillation. The damping by the load has been adjusted so that some time later the energy from the Ls is gone into the load, but the energy in both C_0 peaks at the original value to in turn restore the original voltage so that the voltage across the transistor is zero again and it can be switched on.

With load, frequency, and duty cycle (0.5) as given parameters and the constraint that the voltage is not only restored, but peaks at the original voltage, the four parameters (L , L_0 , C and C_0) are determined. The class E amplifier takes the finite on resistance into account and tries to make the current touch the bottom at zero. This means that the voltage and the current at the transistor are symmetric with respect to time. The Fourier transform allows an elegant formulation to generate the complicated LC networks and says that the first harmonic is passed into the load, all even harmonics are shorted and all higher odd harmonics are open.

Class E uses a significant amount of second-harmonic voltage. The second harmonic can be used to reduce the overlap with edges with finite sharpness. For this to work, energy on the second harmonic has to flow from the load into the transistor, and no source for this is visible in the circuit diagram. In reality, the impedance is mostly reactive and the only reason for it is that class E is a class F amplifier with a much simplified load network and thus has to deal with imperfections.

In many amateur simulations of class E amplifiers, sharp current edges are assumed nullifying the very motivation for class E and measurements near the transit frequency of

the transistors show very symmetric curves, which look much similar to class F simulations.

The class E amplifier was invented in 1972 by Nathan O. Sokal and Alan D. Sokal, and details were first published in 1975. Some earlier reports on this operating class have been published in Russian.

Class F

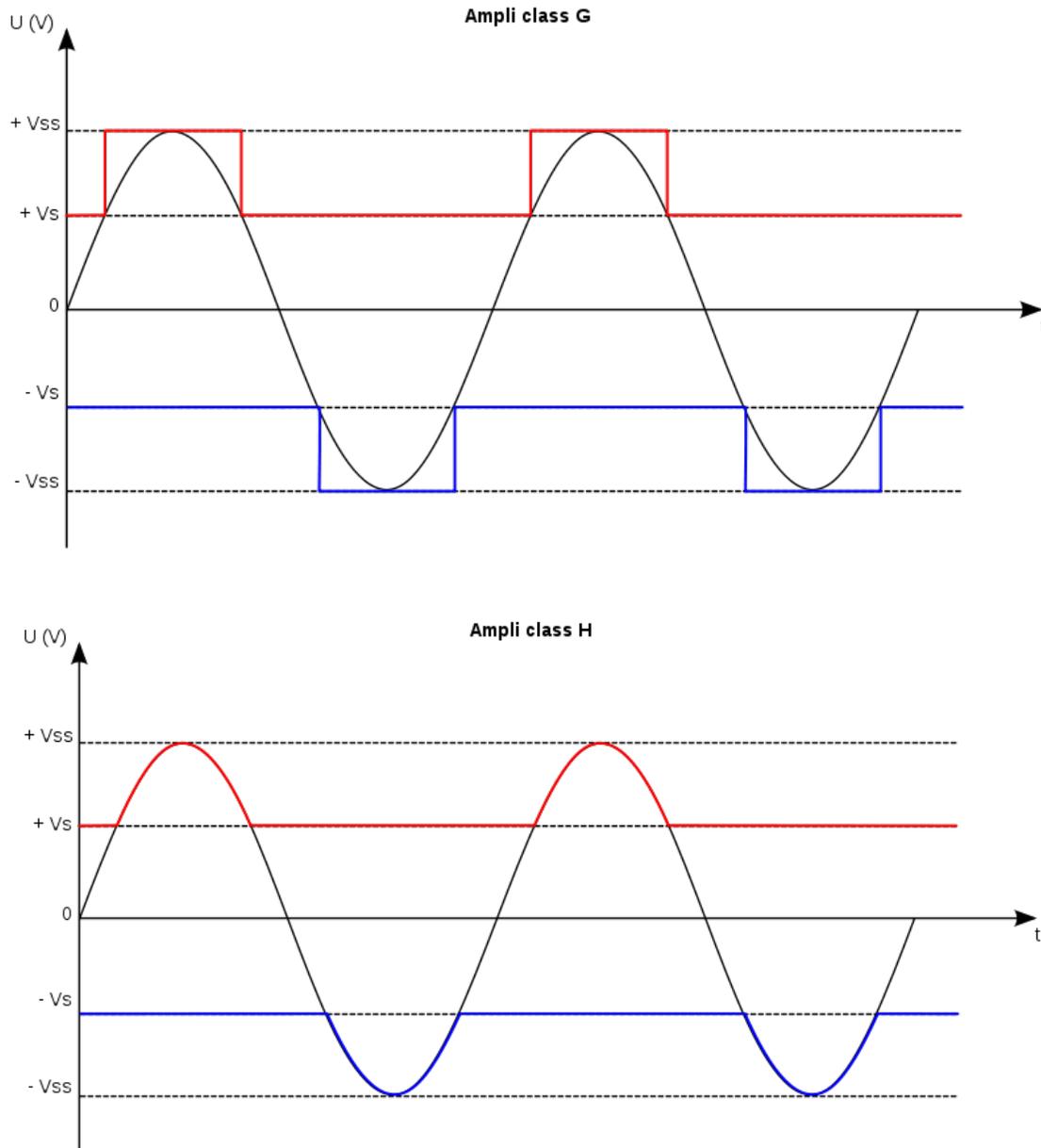
In push-pull amplifiers and in CMOS, the even harmonics of both transistors just cancel. Experiment shows that a square wave can be generated by those amplifiers and theory shows that square waves do consist of odd harmonics only. In a class D amplifier, the output filter blocks all harmonics; i.e., the harmonics see an open load. So even small currents in the harmonics suffice to generate a voltage square wave. The current is in phase with the voltage applied to the filter, but the voltage across the transistors is out of phase. Therefore, there is a minimal overlap between current through the transistors and voltage across the transistors. The sharper the edges, the lower the overlap.

While class D sees the transistors and the load as two separate modules, class F admits imperfections like the parasitics of the transistor and tries to optimise the global system to have a high impedance at the harmonics. Of course there has to be a finite voltage across the transistor to push the current across the on-state resistance. Because the combined current through both transistors is mostly in the first harmonic, it looks like a sine. That means that in the middle of the square the maximum of current has to flow, so it may make sense to have a dip in the square or in other words to allow some overswing of the voltage square wave. A class F load network by definition has to transmit below a cutoff frequency and reflect above.

Any frequency lying below the cutoff and having its second harmonic above the cutoff can be amplified, that is an octave bandwidth. On the other hand, an inductive-capacitive series circuit with a large inductance and a tunable capacitance may be simpler to implement. By reducing the duty cycle below 0.5, the output amplitude can be modulated. The voltage square waveform will degrade, but any overheating is compensated by the lower overall power flowing. Any load mismatch behind the filter can only act on the first harmonic current waveform, clearly only a purely resistive load makes sense, then the lower the resistance, the higher the current.

Class F can be driven by sine or by a square wave, for a sine the input can be tuned by an inductor to increase gain. If class F is implemented with a single transistor, the filter is complicated to short the even harmonics. All previous designs use sharp edges to minimise the overlap.

Classes G and H



There are a variety of amplifier designs that enhance class AB output stages with more efficient techniques to achieve greater efficiencies with low distortion. These designs are common in large audio amplifiers since the heatsinks and power transformers would be prohibitively large (and costly) without the efficiency increases. The terms "class G" and "class H" are used interchangeably to refer to different designs, varying in definition from one manufacturer or paper to another.

Class G amplifiers (which use "rail switching" to decrease power consumption and increase efficiency) are more efficient than class AB amplifiers. These amplifiers provide

several power rails at different voltages and switch between them as the signal output approaches each level. Thus, the amplifier increases efficiency by reducing the wasted power at the output transistors. Class G amplifiers are more efficient than class AB but less efficient when compared to class D, without the negative EMI effects of class D.

Class H amplifiers take the idea of class G one step further creating an infinitely variable supply rail. This is done by modulating the supply rails so that the rails are only a few volts larger than the output signal at any given time. The output stage operates at its maximum efficiency all the time. Switched-mode power supplies can be used to create the tracking rails. Significant efficiency gains can be achieved but with the drawback of more complicated supply design and reduced THD performance.

The voltage signal shown is thus a larger version of the input, but has been changed in sign (inverted) by the amplification. Other arrangements of amplifying device are possible, but that given (that is, common emitter, common source or common cathode) is the easiest to understand and employ in practice. If the amplifying element is linear, then the output will be faithful copy of the input, only larger and inverted. In practice, transistors are not linear, and the output will only approximate the input. Non-linearity from any of several sources is the origin of distortion within an amplifier. Which class of amplifier (A, B, AB or C) depends on how the amplifying device is biased — in the diagrams the bias circuits are omitted for clarity.

Any real amplifier is an imperfect realization of an ideal amplifier. One important limitation of a real amplifier is that the output it can generate is ultimately limited by the power available from the power supply. An amplifier will saturate and clip the output if the input signal becomes too large for the amplifier to reproduce or if operational limits for a device are exceeded.

Doherty amplifiers

A hybrid configuration receiving new attention is the Doherty amplifier, invented in 1934 by William H. Doherty for Bell Laboratories (whose sister company, Western Electric, was then an important manufacturer of radio transmitters). The Doherty amplifier consists of a class B *primary* or *carrier* stage in parallel with a class C *auxiliary* or *peak* stage. The input signal is split to drive the two amplifiers and a combining network sums the two output signals. Phase shifting networks are employed in the inputs and the outputs. During periods of low signal level, the class B amplifier efficiently operates on the signal and the class C amplifier is cutoff and consumes little power. During periods of high signal level, the class B amplifier delivers its maximum power and the class C amplifier delivers up to its maximum power. The efficiency of previous AM transmitter designs was proportional to modulation but, with average modulation typically around 20%, transmitters were limited to less than 50% efficiency. In Doherty's design, even with zero modulation, a transmitter could achieve at least 60% efficiency.

As a successor to Western Electric for broadcast transmitters, the Doherty concept was considerably refined by Continental Electronics Manufacturing Company of Dallas, TX.

Perhaps, the ultimate refinement was the screen-grid modulation scheme invented by Joseph B. Sinton. The Sinton amplifier consists of a class C primary or carrier stage in parallel with a class C auxiliary or peak stage. The stages are split and combined through 90-degree phase shifting networks as in the Doherty amplifier. The unmodulated radio frequency carrier is applied to the control grids of both tubes. Carrier modulation is applied to the screen grids of both tubes. The bias point of the carrier and peak tubes is different, and is established such that the peak tube is cutoff when modulation is absent (and the amplifier is producing rated unmodulated carrier power) whereas both tubes contribute twice the rated carrier power during 100% modulation (as four times the carrier power is required to achieve 100% modulation). As both tubes operate in class C, a significant improvement in efficiency is thereby achieved in the final stage. In addition, as the tetrode carrier and peak tubes require very little drive power, a significant improvement in efficiency within the driver stage is achieved as well (317C, et al.). The released version of the Sinton amplifier employs a cathode-follower modulator, not a push-pull modulator. Previous Continental Electronics designs, by James O. Weldon and others, retained most of the characteristics of the Doherty amplifier but added screen-grid modulation of the driver (317B, et al.).

The Doherty amplifier remains in use in very-high-power AM transmitters, but for lower-power AM transmitters, vacuum-tube amplifiers in general were eclipsed in the 1980s by arrays of solid-state amplifiers, which could be switched on and off with much finer granularity in response to the requirements of the input audio. However, interest in the Doherty configuration has been revived by cellular-telephone and wireless-Internet applications where the sum of several constant-envelope users creates an aggregate AM result. The main challenge of the Doherty amplifier for digital transmission modes is in aligning the two stages and getting the class-C amplifier to turn on and off very quickly.

Recently, Doherty amplifiers have found widespread use in cellular base station transmitters for GHz frequencies. Implementations for transmitters in mobile devices have also been demonstrated.

Special classes

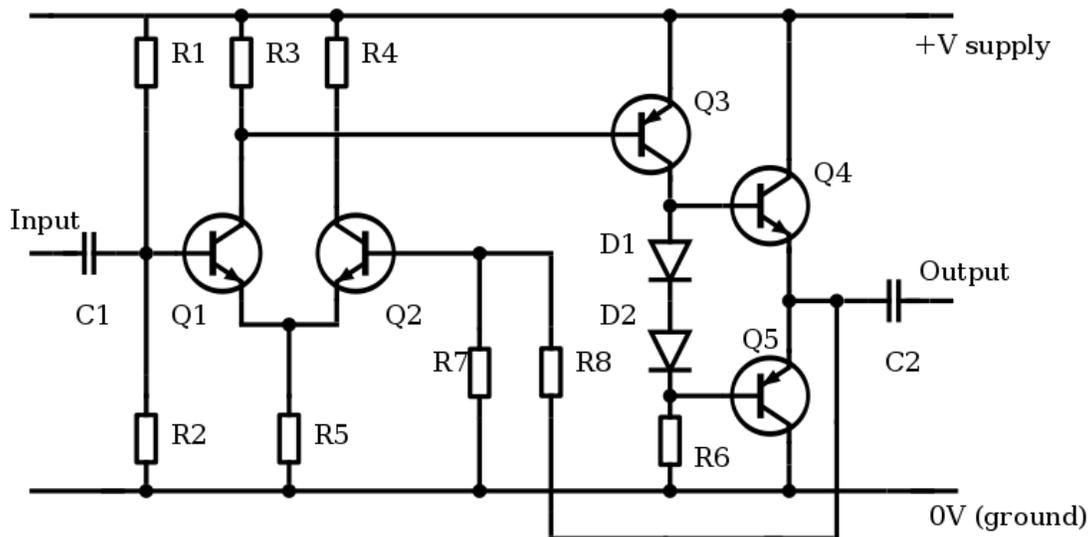
Several audio amplifier manufacturers have started "inventing" new classes as a way to differentiate themselves. These class names usually do not reflect any revolutionary amplification technique, and are used mostly for marketing purposes. This can easily be determined by the fact that the class name is trademarked or copyrighted. For example, Crown's K and I-Tech Series as well as several other models utilise Crown's patented class I (or BCA) technology. Lab.gruppen use a form of class D amplifier called class TD or tracked class D which tracks the waveform to more accurately amplify it without the drawbacks of traditional class D amplifiers.

"Class T" was a trademark of TriPath company which manufactures audio amplifier ICs. This new class T is a revision of the common class D amplifier, but with changes to ensure fidelity over the full audio spectrum, unlike traditional class D designs. It operates at different frequencies depending on the power output, with values ranging from as low

as 200 kHz to 1.2 MHz, using a proprietary modulator. Tripath ceased operations in 2007, its patents acquired by Cirrus Logic for their Mixed-Signal Audio division. Some Kenwood Recorder use Class W amplifier

"Class Z" is a trademark of Zetex Semiconductors (now part of Diodes Inc. of Dallas, TX) and is a direct-digital-feedback technology. Zetex-patented circuits are being utilised in the latest power amplifiers by NAD Electronics of Canada.

Amplifier circuit



The practical amplifier circuit to the right could be the basis for a moderate-power audio amplifier. It features a typical (though substantially simplified) design as found in modern amplifiers, with a class AB push-pull output stage, and uses some overall negative feedback. Bipolar transistors are shown, but this design would also be realizable with FETs or valves.

The input signal is coupled through capacitor C1 to the base of transistor Q1. The capacitor allows the AC signal to pass, but blocks the DC bias voltage established by resistors R1 and R2 so that any preceding circuit is not affected by it. Q1 and Q2 form a differential amplifier (an amplifier that multiplies the difference between two inputs by some constant), in an arrangement known as a long-tailed pair. This arrangement is used to conveniently allow the use of negative feedback, which is fed from the output to Q2 via R7 and R8.

The negative feedback into the difference amplifier allows the amplifier to compare the input to the actual output. The amplified signal from Q1 is directly fed to the second stage, Q3, which is a common emitter stage that provides further amplification of the signal and the DC bias for the output stages, Q4 and Q5. R6 provides the load for Q3 (A better design would probably use some form of active load here, such as a constant-current sink). So far, all of the amplifier is operating in class A. The output pair are

arranged in class AB push–pull, also called a complementary pair. They provide the majority of the current amplification (while consuming low quiescent current) and directly drive the load, connected via DC-blocking capacitor C2. The diodes D1 and D2 provide a small amount of constant voltage bias for the output pair, just biasing them into the conducting state so that crossover distortion is minimized. That is, the diodes push the output stage firmly into class-AB mode (assuming that the base-emitter drop of the output transistors is reduced by heat dissipation).

This design is simple, but a good basis for a practical design because it automatically stabilises its operating point, since feedback internally operates from DC up through the audio range and beyond. Further circuit elements would probably be found in a real design that would roll off the frequency response above the needed range to prevent the possibility of unwanted oscillation. Also, the use of fixed diode bias as shown here can cause problems if the diodes are not both electrically and thermally matched to the output transistors — if the output transistors turn on too much, they can easily overheat and destroy themselves, as the full current from the power supply is not limited at this stage.

A common solution to help stabilise the output devices is to include some emitter resistors, typically an ohm or so. Calculating the values of the circuit's resistors and capacitors is done based on the components employed and the intended use of the amp.

Notes on implementation

Real world amplifiers are imperfect.

- One consequence is that the power supply itself may influence the output, and must itself be considered when designing the amplifier
- The amplifier circuit has an "open loop" performance, that can be described as various parameters (gain, slew rate, output impedance, distortion, bandwidth, signal to noise ratio, etc.)
- Many modern amplifiers use negative feedback techniques to hold the gain at the desired value.

Different methods of supplying power result in many different methods of bias. Bias is a technique by which the active devices are set up to operate in a particular regime, or by which the DC component of the output signal is set to the midpoint between the maximum voltages available from the power supply. Most amplifiers use several devices at each stage; they are typically matched in specifications except for polarity. Matched inverted polarity devices are called complementary pairs. Class A amplifiers generally use only one device, unless the power supply is set to provide both positive and negative voltages, in which case a dual device symmetrical design may be used. Class C amplifiers, by definition, use a single polarity supply.

Amplifiers often have multiple stages in cascade to increase gain. Each stage of these designs may be a different type of amp to suit the needs of that stage. For instance, the first stage might be a class A stage, feeding a class AB push–pull second stage, which

then drives a class G final output stage, taking advantage of the strengths of each type, while minimizing their weaknesses.

Chapter 6

Transmitter



Antenna tower of Crystal Palace transmitter, London

In electronics and telecommunications a **transmitter** or **radio transmitter** is an electronic device which, with the aid of an antenna, produces radio waves. The transmitter itself generates a radio frequency alternating current, which is applied to the antenna. When excited by this alternating current, the antenna radiates radio waves. In addition to their use in broadcasting, transmitters are necessary component parts of many electronic devices that communicate by radio, such as cell phones, Wifi and Bluetooth enabled devices, garage door openers, two-way radios in aircraft, ships, and spacecraft, radar sets, and navigational beacons. The term *transmitter* is usually limited to equipment that generates radio waves for communication purposes; or radiolocation, such as radar and navigational transmitters. Generators of radio waves for heating or industrial purposes, such as microwave ovens or diathermy equipment, are not usually called transmitters even though they often have similar circuits.

The term is popularly used more specifically to refer to transmitting equipment used for broadcasting, as in radio transmitter or television transmitter. This usage usually includes both the transmitter proper as described above, and the antenna, and often the building it is housed in.

An unrelated use of the term is in industrial process control, where a "transmitter" is a device which converts measurements from a sensor into a signal, and sends it, usually via wires, to be received by some display or control device located a distance away.

Description

A transmitter can be a separate piece of electronic equipment, or an electrical circuit within another electronic device. A transmitter and receiver combined in one unit is called a transceiver. The term transmitter is often abbreviated "XMTR" or "TX" in technical documents. The purpose of most transmitters is radio communication of information over a distance. The information is provided to the transmitter in the form of an electronic signal, such as an audio (sound) signal from a microphone, a video (TV) signal from a TV camera, or in wireless networking devices a digital signal from a computer. The transmitter combines the information signal to be carried with the radio frequency signal which generates the radio waves, which is often called the carrier. This process is called *modulation*. The information can be added to the carrier in several different ways, in different types of transmitter. In an amplitude modulation (AM) transmitter, the information is added to the radio signal by varying its amplitude (strength). In a frequency modulation (FM) transmitter, it is added by varying the radio signal's frequency slightly. Many other types of modulation are used.

Legal restrictions

In most parts of the world, use of transmitters is strictly controlled by law because of the potential for dangerous interference with other radio transmissions (for example to emergency communications). Transmitters must be licensed by governments, under a variety of license classes depending on use: (broadcast, marine radio, Airband, Amateur etc.), and are restricted to certain frequencies and power levels. In some classes each

transmitter is given a unique call sign consisting of a string of letters and numbers which must be used as an identifier in transmissions. The operator of the transmitter usually must hold a government license, such as a general radiotelephone operator license, which is obtained by passing a test demonstrating adequate technical and legal knowledge of safe radio operation.

An exception is made allowing the unlicensed use of low-power short-range transmitters in devices such as wireless microphones, cordless telephones, walkie-talkies, Wifi and Bluetooth, garage door openers, and baby monitors. In the US, these fall under Part 15 of the Federal Communications Commission (FCC) regulations. Although they can be operated without a license, these devices still generally must be type-approved before sale.

How it works

A radio transmitter is an electronic circuit which transforms electric power from a battery or electrical mains into a radio frequency alternating current, which reverses direction millions to billions of times per second. The energy in such a rapidly-reversing current can radiate off a conductor (the antenna) as electromagnetic waves (radio waves). The transmitter also "piggybacks" information, such as an audio or video signal, onto the radio frequency current to be carried by the radio waves. When they strike the antenna of a radio receiver, the waves excite similar (but less powerful) radio frequency currents in it. The radio receiver extracts the information from the received waves. A practical radio transmitter usually consists of these parts:

- A power supply circuit to transform the input electrical power to the higher voltages needed to produce the required power output.
- An electronic oscillator circuit to generate the radio frequency signal. This usually generates a sine wave of constant amplitude often called the carrier wave. In most modern transmitters this is a crystal oscillator in which the frequency is precisely controlled by the vibrations of a quartz crystal.
- A modulator circuit to add the information to be transmitted to the carrier wave produced by the oscillator. This is done by varying some aspect of the carrier wave. The information is provided to the transmitter either in the form of an audio signal, which represents sound, a video signal, or for data in the form of a binary digital signal.
 - In an AM (amplitude modulation) transmitter the amplitude (strength) of the carrier wave is varied in proportion to the audio signal.
 - In an FM (frequency modulation) transmitter the frequency of the carrier is varied by the audio signal.
 - In an FSK (frequency-shift keying) transmitter, which transmits digital data, the frequency of the carrier is shifted between two frequencies which represent the two binary digits, 0 and 1.

Many other types of modulation are also used. In large transmitters the oscillator and modulator together are often referred to as the *exciter*.

- An RF power amplifier to increase the power of the signal, to increase the range of the radio waves.
- An impedance matching (antenna tuner) circuit to match the impedance of the transmitter to the impedance of the antenna (or the transmission line to the antenna), to transfer power efficiently to the antenna. If these impedances are not equal, it causes a condition called standing waves, in which the power is reflected back from the antenna toward the transmitter, wasting power and sometimes overheating the transmitter.

In higher frequency transmitters, in the UHF and microwave range, oscillators that operate stably at the output frequency cannot be built. In these transmitters the oscillator usually operates at a lower frequency, usually a submultiple of the output frequency, and this intermediate frequency (IF) is multiplied to get a signal at the output frequency by frequency multipliers.

History

The first primitive radio transmitters (called Hertzian oscillators) were built by German physicist Heinrich Hertz in 1887 during his pioneering investigations of radio waves. These generated radio waves by a high voltage spark between two conductors. These spark-gap transmitters were used during the first three decades of radio (1887-1917), called the wireless telegraphy era. Short-lived competing techniques came into use after the turn of the century, such as the Alexanderson alternator and Poulsen Arc transmitters. But all these early technologies were replaced by vacuum tube transmitters in the 1920s, because they were inexpensive and produced continuous waves, which could be modulated to transmit audio (sound) using amplitude modulation (AM) and frequency modulation (FM). This made possible commercial radio broadcasting, which began about 1920. The development of radar before and during World War 2 was a great stimulus to the evolution of high frequency transmitters in the UHF and microwave ranges, using new devices such as the magnetron and traveling wave tube. In recent years, the need to conserve crowded radio spectrum bandwidth has driven the development of new types of transmitters such as spread spectrum.

Broadcast transmitters



Commercial FM broadcasting transmitter at radio station WDET-FM, Wayne State University, Detroit, USA. It broadcasts at 101.9 MHz with a radiated power of 48 kW.

Frequency Control

Power output

In broadcasting and telecommunication, the part which contains the oscillator, modulator, and sometimes audio processor, is called the "exciter". Most transmitters use heterodyne principle, so they also have a frequency conversion units. Confusingly, the high-power amplifier which the exciter then feeds into is often called the "transmitter" by broadcast engineers. The final output is given as transmitter power output (TPO), although this is not what most stations are rated by.

Effective radiated power (ERP) is used when calculating station coverage, even for most non-broadcast stations. It is the TPO, minus any attenuation or radiated loss in the line to the antenna, multiplied by the gain (magnification) which the antenna provides toward the horizon. This antenna gain is important, because achieving a desired signal strength without it would result in an enormous electric utility bill for the transmitter, and a prohibitively expensive transmitter. For most large stations in the VHF- and UHF-range, the transmitter power is no more than 20% of the ERP.

For VLF, LF, MF and HF the ERP is typically not determined separately. In most cases the transmission power found in lists of transmitters is the value for the output of the transmitter. This is only correct for omnidirectional aerials with a length of a quarter wavelength or shorter. For other aerial types there are gain factors, which can reach values until 50 for shortwave directional beams in the direction of maximum beam intensity.

Since some authors take account of gain factors of aerials of transmitters for frequencies below 30 MHz and others not, there are often discrepancies of the values of transmitted powers.

Power supply

Transmitters are sometimes fed from a higher voltage level of the power supply grid than necessary in order to improve security of supply. For example, the Allouis, Konstantynow and Roumoules transmitters are fed from the high-voltage network (110 kV in Alouis and Konstantynow, 150 kV in Roumoules) even though a power supply from the medium-voltage level of the power grid (about 20 kV) would be able to deliver enough power.

Cooling of final stages

Low-power transmitters do not require special cooling equipment. Modern transmitters can be incredibly efficient, with efficiencies exceeding 98 percent. However, a broadcast transmitter with a megawatt power stage transferring 98% of that into the antenna can also be viewed as a 20 kilowatt electric heater.

For medium-power transmitters, up to a few hundred watts, air cooling with fans is used. At power levels over a few kilowatts, the output stage is cooled by a forced liquid cooling system analogous to an automobile cooling system. Since the coolant directly touches the high-voltage anodes of the tubes, only distilled, deionised water or a special dielectric coolant can be used in the cooling circuit. This high-purity coolant is in turn cooled by a heat exchanger, where the second cooling circuit can use water of ordinary quality because it is not in contact with energized parts. Very-high-power tubes of small physical size may use evaporative cooling by water in contact with the anode. The production of steam allows a high heat flow in a small space.

Protection equipment

The high voltages used in high power transmitters (up to 40 kV) require extensive protection equipment. Also, transmitters are exposed to damage from lightning. Transmitters may be damaged if operated without an antenna, so protection circuits must detect the loss of the antenna and switch off the transmitter immediately. Tube-based transmitters must have power applied in the proper sequence, with the filament voltage applied before the anode voltage, otherwise the tubes can be damaged. The output stage must be monitored for standing waves, which indicate that generated power is not being radiated but instead is being reflected back into the transmitter.

Lightning protection is required between the transmitter and antenna. This consists of spark gaps and gas-filled surge arresters to limit the voltage that appears on the transmitter terminals. The control instrument that measures the voltage standing-wave ratio switches the transmitter off briefly if a higher voltage standing-wave ratio is detected after a lightning strike, as the reflections are probably due to lightning damage. If this does not succeed after several attempts, the antenna may be damaged and the transmitter should remain switched off. In some transmitting plants UV detectors are fitted in critical places, to switch off the transmitter if an arc is detected. The operating voltages, modulation factor, frequency and other transmitter parameters are monitored for protection and diagnostic purposes, and may be displayed locally and/or at a remote control room.

Building

A commercial transmitter site will usually have a control building to shelter the transmitter components and control devices. This is usually a purely functional building, which may contain apparatus for both radio and television transmitters. To reduce transmission line loss the transmitter building is usually immediately adjacent to the antenna for VHF and UHF sites, but for lower frequencies it may be desirable to have a distance of a few score or several hundred metres between the building and the antenna. Some transmitting towers have enclosures built into the tower to house radio relay link transmitters or other, relatively low-power transmitters. A few transmitter buildings may include limited broadcasting facilities to allow a station to use the building as a backup studio in case of incapacitation of the main facility.

Legal and regulatory aspects

Since radio waves go over borders, international agreements control radio transmissions. In European countries like Germany often the national Post Office is the regulating authority. In the United States broadcast and industrial transmitters are regulated by the Federal Communications Commission (FCC). In Canada technical aspects of broadcast and radio transmitters are controlled by Industry Canada, but broadcast content is regulated separately by the Canadian Radio-television and Telecommunications Commission (CRTC). In Australia transmitters, spectrum, and content are controlled by the Australian Communications and Media Authority (ACMA). The International

Telecommunication Union (ITU) helps managing the radio-frequency spectrum internationally.

Planning

As in any costly project, the planning of a high power transmitter site requires great care. This begins with the location. A minimum distance, which depends on the transmitter frequency, transmitter power, and the design of the transmitting antennas, is required to protect people from the radio frequency energy. Antenna towers are often very tall and therefore flight paths must be evaluated. Sufficient electric power must be available for high power transmitters. Transmitters for long and medium wave require good grounding and soil of high electrical conductivity. Locations at the sea or in river valleys are ideal, but the flood danger must be considered. Transmitters for UHF are best on high mountains to improve the range. The antenna pattern must be considered because it is costly to change the pattern of a long-wave or medium-wave antenna.



Antenna guyed tower

Transmitting antennas for long and medium wave are usually implemented as a mast radiator. Similar antennas with smaller dimensions are used also for short wave transmitters, if these send in the round spray enterprise. For arranging radiation at free standing steel towers fastened planar arrays are used. Radio towers for UHF and TV transmitters can be implemented in principle as grounded constructions. Towers may be steel lattice masts or reinforced concrete towers with antennas mounted at the top. Some transmitting towers for UHF have high-altitude operating rooms and/or facilities such as restaurants and observation platforms, which are accessible by elevator. Such towers are usually called TV tower. For microwaves one frequently uses parabolic antennas. These can be set up for applications of radio relay links on transmitting towers for FM to special platforms. For example, large parabolic antennas ranging from 3 to 100 meters in diameter are necessary to pass on signals to television satellites and space vehicles. These plants, which can be used if necessary also as radio telescope, are established on free standing constructions, whereby there are also numerous special designs, like the radio telescope in Arecibo.

Just as important as the planning of the construction and location of the transmitter is how its output fits in with existing transmissions. Two transmitters cannot broadcast on the same frequency in the same area as this would cause co-channel interference. For a good example of how the channel planners have dovetailed different transmitters' outputs see Crystal Palace UHF TV channel allocations. This reference also provides a good example of a grouped transmitter, in this case an A group. That is, all of its output is within the bottom third of the UK UHF television broadcast band. The other two groups (B and C/D) utilise the middle and top third of the band. By replicating this grouping across the country (using different groups for adjacent transmitters), co-channel interference can be minimised, and in addition, those in marginal reception areas can use more efficient grouped receiving antennas. Unfortunately, in the UK, this carefully planned system has had to be compromised with the advent of digital broadcasting which (during the changeover period at least) requires yet more channel space, and consequently the additional digital broadcast channels cannot always be fitted within the transmitter's existing group. Thus many UK transmitters have become "wideband" with the consequent need for replacement of receiving antennas. Once the Digital Switch Over (DSO) occurs the plan is that most transmitters will revert to their original groups, source Ofcom July 2007 .

Further complication arises when adjacent transmitters have to transmit on the same frequency and under these circumstances the broadcast radiation patterns are attenuated in the relevant direction(s). A good example of this is in the United Kingdom, where the Waltham transmitting station broadcasts at high power on the same frequencies as the Sandy Heath transmitting station's high power transmissions, with the two being only 50 miles apart. Thus Waltham's antenna array does not broadcast these two channels in the direction of Sandy Heath and vice versa.

Where a particular service needs to have wide coverage, this is usually achieved by using multiple transmitters at different locations. Usually, these transmitters will operate at different frequencies to avoid interference where coverage overlaps. Examples include

national broadcasting networks and cellular networks. In the latter, frequency switching is automatically done by the receiver as necessary, in the former, manual retuning is more common (though the Radio Data System is an example of automatic frequency switching in broadcast networks). Another system for extending coverage using multiple transmitters is quasi-synchronous transmission, but this is rarely used nowadays.

Main and relay (repeater) transmitters

Transmitting stations are usually either classified as main stations or relay stations (also known as repeaters, translators or sometimes "transposers".)

Main stations are defined as those that generate their own modulated output signal from a baseband (unmodulated) input. Usually main stations operate at high power and cover large areas.

Relay stations (translators) take an already modulated input signal, usually by direct reception of a parent station off the air, and simply rebroadcast it on another frequency. Usually relay stations operate at medium or low power, and are used to fill in pockets of poor reception within, or at the fringe of, the service area of a parent main station.

Note that a main station may also take its input signal directly off-air from another station, however this signal would be fully demodulated to baseband first, processed, and then remodulated for transmission.

Transmitters in culture

Some cities in Europe, like Mühlacker, Ismaning, Langenberg, Kalundborg, Hörby and Allouis became famous as sites of powerful transmitters. For example, Goliath transmitter was a VLF transmitter of the German Navy during World War II located near Kalbe an der Milde in Saxony-Anhalt, Germany. Some transmitting towers like the radio tower Berlin or the TV tower Stuttgart have become landmarks of cities. Many transmitting plants have very high radio towers that are masterpieces of engineering.

Having the tallest building in the world, the nation, the state/province/prefecture, city, etc., has often been considered something to brag about. Often, builders of high-rise buildings have used transmitter antennas to lay claim to having the tallest building. A historic example was the "tallest building" feud between the Chrysler Building and the Empire State Building in New York, New York.

Some towers have an observation deck accessible to tourists. An example is the Ostankino Tower in Moscow, which was completed in 1967 on the 50th anniversary of the October Revolution to demonstrate the technical abilities of the Soviet Union. As very tall radio towers of any construction type are prominent landmarks, requiring careful planning and construction, and high-power transmitters especially in the long- and medium-wave ranges can be received over long distances, such facilities were often

mentioned in propaganda. Other examples were the Deutschlandsender Herzberg/Elster and the Warsaw Radio Mast.

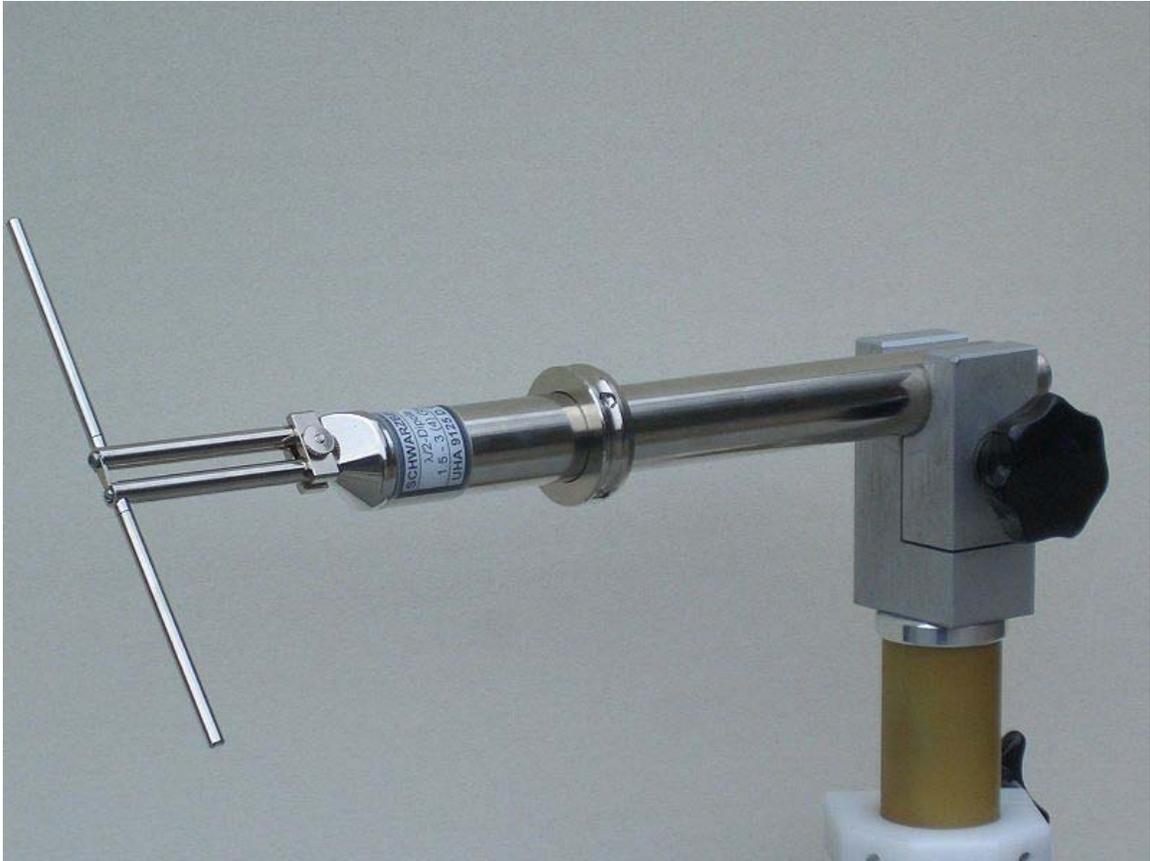
KVLY-TV's tower located near Blanchard, North Dakota was the tallest artificial structure in the world when it was completed in 1963. It was surpassed in 1974 by the Warszawa radio mast, but regained its title when the latter collapsed in 1991. It was surpassed by the Burj Khalifa skyscraper in early 2009, but the KVLY-TV mast is still the tallest transmitter.

Chapter 7

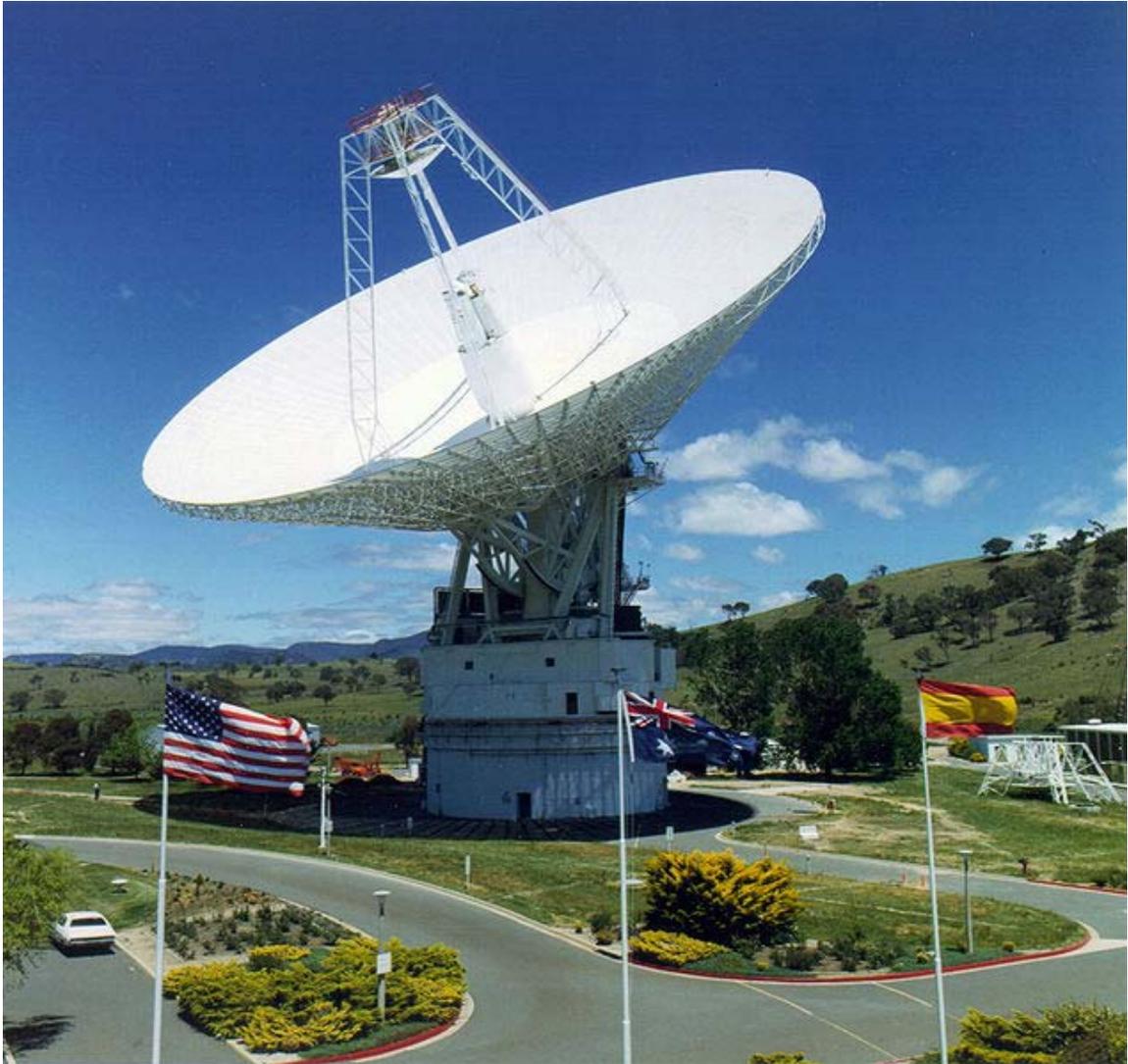
Antenna



Whip antenna on car



Half-wave dipole antenna



Large parabolic antenna for communicating with spacecraft



Rooftop directional antennas, typical for use at VHF and UHF frequencies

An **antenna** (or **aerial**) is a transducer that transmits or receives electromagnetic waves. In other words, antennas convert electromagnetic radiation into an electrical signal and vice versa. They are used to transmit and receive electromagnetic radiation of radio frequency, that is, radio waves, and are essential to the operation of all radio equipment. Antennas are used in systems such as radio and television broadcasting, point-to-point radio communication, wireless LAN, cell phones, radar, and spacecraft communication.

Physically, an antenna is an arrangement of one or more conductors, usually called *elements* in this context. In transmission, an alternating current is created in the elements by applying a voltage at the antenna terminals, causing the elements to radiate an electromagnetic field. In reception, the inverse occurs: an electromagnetic field from

another source induces an alternating current in the elements and a corresponding voltage at the antenna's terminals. In a parabolic antenna, a feed antenna of that sort is augmented by a much larger curved reflecting surface which creates a beam of radio waves similar to that of a searchlight. Since they are much larger than a wavelength, the parabolic and other so-called *aperture antennas* are widely used only at higher, especially microwave frequencies.

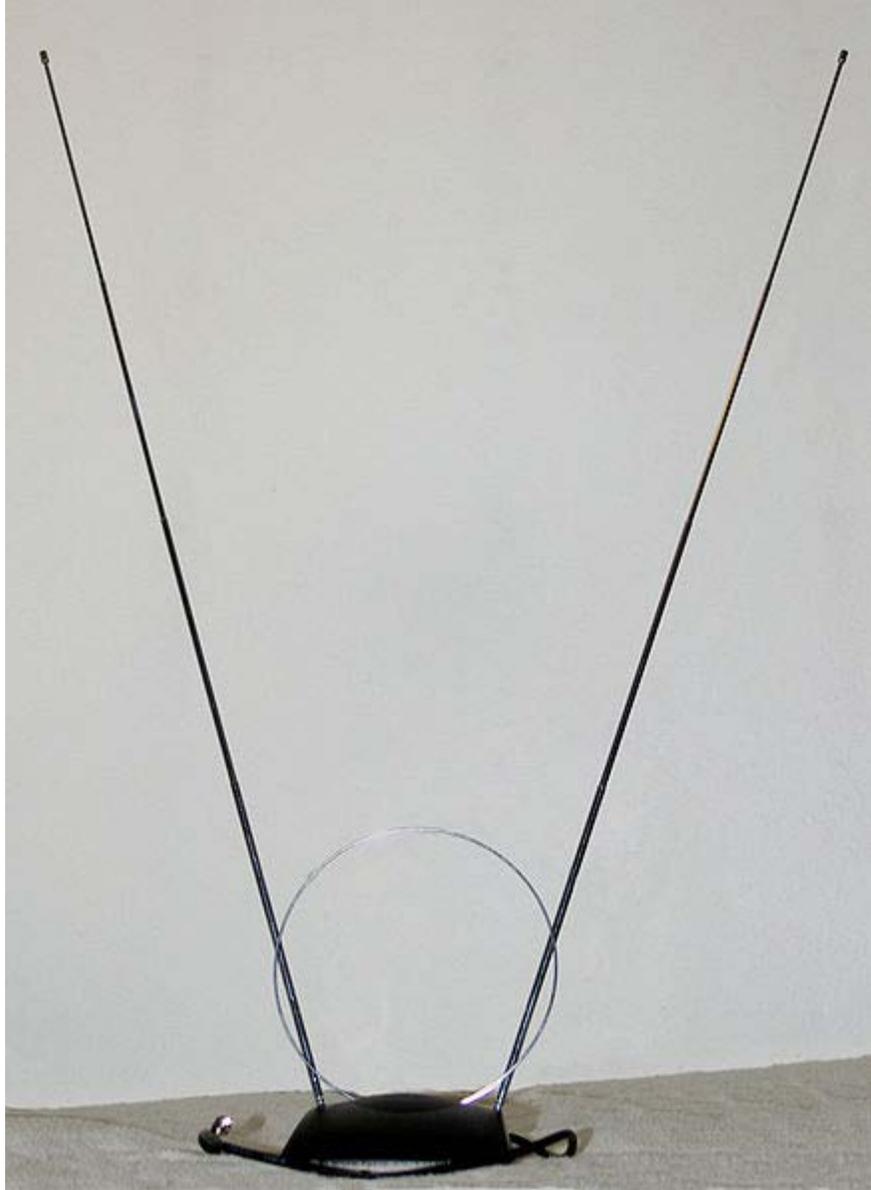
The first antennas were built in 1888 by Heinrich Hertz (1857–1894) in his pioneering experiments to prove the existence of electromagnetic waves predicted by the theory of James Clerk Maxwell. Hertz placed dipole antennas at the focal point of parabolic reflectors for both transmitting and receiving. He published his work and installation drawings in *Annalen der Physik und Chemie* (vol. 36, 1889).

Terminology

The words *antenna* (plural: *antennas*) and *aerial* are used interchangeably; but usually a rigid metallic structure is termed an antenna and a wire format is called an aerial. In the United Kingdom and other British English speaking areas the term aerial is more common, even for rigid types. The noun *aerial* is occasionally written with a diaeresis mark—*aërial*—in recognition of the original spelling of the adjective *aërial* from which the noun is derived.

The origin of the word *antenna* relative to wireless apparatus is attributed to Guglielmo Marconi. In 1895, while testing early radio apparatuses in the Swiss Alps at Salvan, Switzerland in the Mont Blanc region, Marconi experimented with early wireless equipment. A 2.5 meter long pole, along which was carried a wire, was used as a radiating and receiving aerial element. In Italian a tent pole is known as *l'antenna centrale*, and the pole with a wire alongside it used as an aerial was simply called *l'antenna*. Until then wireless radiating transmitting and receiving elements were known simply as aërials or terminals. Marconi's use of the word *antenna* (Italian for *pole*) would become a popular term for what today is uniformly known as the *antenna*.

In common usage, the word *antenna* may refer broadly to an entire assembly including support structure, enclosure (if any), etc. in addition to the actual functional components. Especially at microwave frequencies, a receiving antenna may include not only the actual electrical antenna but an integrated preamplifier and/or mixer.



"Rabbit ears" dipole antenna for television reception



Cell phone base station antennas



Satellite link antenna used by Himalaya Television Nepal



Bundesarchiv, Bild 183-20802-0001
Foto: Braun I, April 1955

Yagi antenna used for mobile military communications station, Dresden,



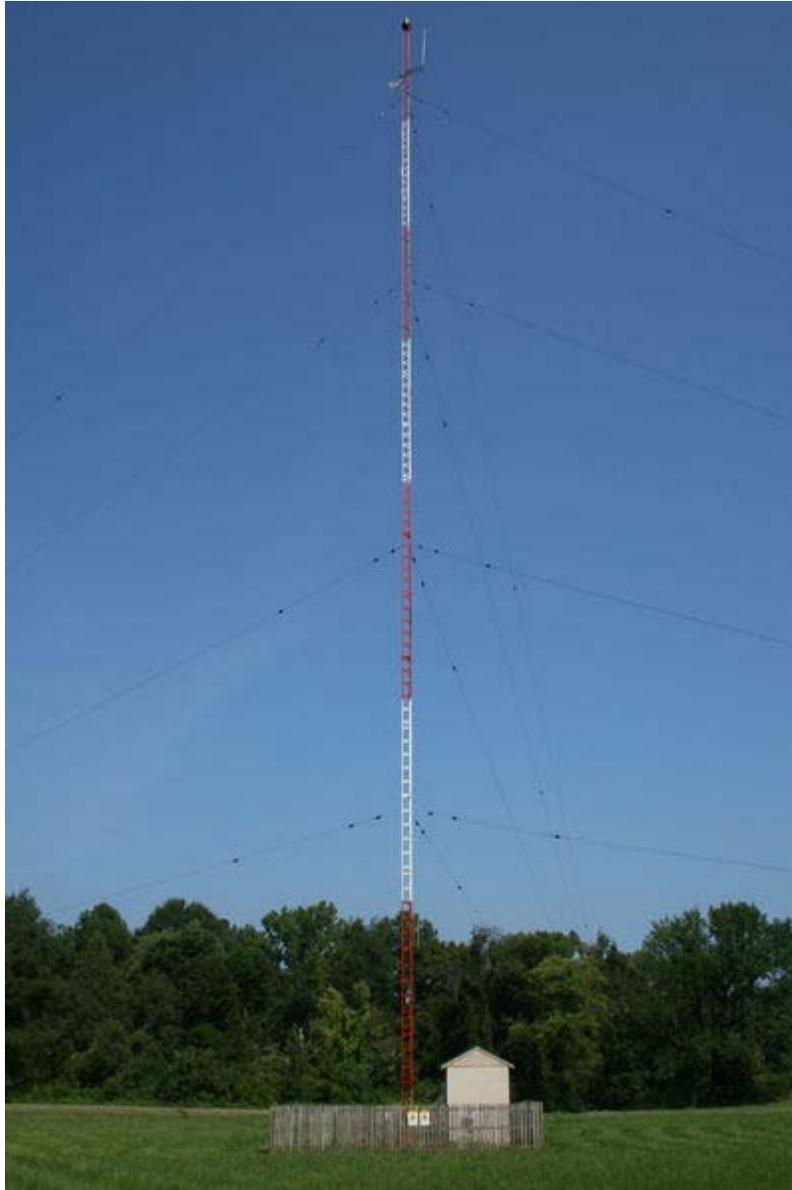
"Super Turnstile" type transmitting antenna for VHF low band television broadcasting station,



Folded dipole antenna



Large Yagi antenna used by amateur radio hobbyist



A vertical mast radiator, Chapel Hill, North Carolina

Overview

Antennas are required by any radio receiver or transmitter in order to couple its electrical connection to the electromagnetic field. Radio waves are electromagnetic waves which carry signals through the air (or through space) at almost the speed of light with almost no transmission loss. Radio transmitters and receivers are used to convey signals (information) in systems including broadcast (audio) radio, television, mobile telephones, wi-fi (WLAN) data networks, trunk lines and point-to-point communications links (telephone, data networks), satellite links, many remote controlled devices such as garage door openers, wireless remote sensors, among many others. Radio waves are also used directly for measurements in technologies including RADAR, GPS, and radio astronomy.

In each and every case, the transmitters and receivers involved require antennas, although these are sometimes hidden (such as the antenna inside an AM radio or inside a laptop computer equipped with wi-fi).

According to their applications and technology available, antennas generally fall in one of two categories:

1. Omni-directional or only weakly directional antennas which receive or radiate more or less in all directions. These are employed when the relative position of the other station is unknown or arbitrary. They are also used at lower frequencies where a directional antenna would be too large, or simply to cut costs in applications where a directional antenna isn't required.
2. Directional or *beam* antennas which are intended to preferentially radiate or receive in a particular direction or directional pattern.

In common usage "omnidirectional" usually refers to all horizontal directions, typically with reduced performance in the direction of the sky or the ground (a truly isotropic radiator is not even possible). A "directional" antenna usually is intended to maximize its coupling to the electromagnetic field in the direction of the other station, or sometimes to cover a particular sector such as a 120° horizontal fan pattern in the case of a panel antenna at a cell site.

One example of omnidirectional antennas is the very common *vertical antenna* or whip antenna consisting of a metal rod (often, but not always, a quarter of a wavelength long). A dipole antenna is similar but consists of two such conductors extending in opposite directions, with a total length that is often, but not always, a half of a wavelength long. Dipoles are typically oriented horizontally in which case they are weakly directional: signals are reasonably well radiated toward or received from all directions with the exception of the direction along the conductor itself; this region is called the antenna blind cone or null.

Both the vertical and dipole antennas are simple in construction and relatively inexpensive. The dipole antenna, which is the basis for most antenna designs, is a balanced component, with equal but opposite voltages and currents applied at its two terminals through a balanced transmission line (or to a coaxial transmission line through a so-called balun). The vertical antenna, on the other hand, is a *monopole* antenna. It is typically connected to the inner conductor of a coaxial transmission line (or a matching network); the shield of the transmission line is connected to ground. In this way, the ground (or any large conductive surface) plays the role of the second conductor of a dipole, thereby forming a complete circuit. Since monopole antennas rely on a conductive ground, a so-called grounding structure may be employed in order to provide a better ground contact to the earth or which itself acts as a ground plane to perform that function regardless of (or in absence of) an actual contact with the earth.

Antennas fancier than the dipole or vertical designs are usually intended to increase the directivity and consequently the gain of the antenna. This can be accomplished in many

different ways leading to a plethora of antenna designs. The vast majority of designs are fed with a balanced line (unlike a monopole antenna) and are based on the dipole antenna with additional components (or *elements*) which increase its directionality.

For instance, a phased array consists of two or more simple antennas which are connected together through an electrical network. This often involves a number of parallel dipole antennas with a certain spacing. Depending on the relative phase introduced by the network, the same combination of dipole antennas can operate as a "broadside array" (directional normal to a line connecting the elements) or as an "end-fire array" (directional along the line connecting the elements). Antenna arrays may employ any basic (omnidirectional or weakly directional) antenna type, such as dipole, loop or slot antennas. These elements are often identical.

However a log-periodic dipole array consists of a number of dipole elements of *different* lengths in order to obtain a somewhat directional antenna having an extremely wide bandwidth: these are frequently used for television reception in fringe areas. The dipole antennas composing it are all considered "active elements" since they are all electrically connected together (and to the transmission line). On the other hand, a superficially similar dipole array, the Yagi-Uda Antenna (or simply "Yagi"), has only one dipole element with an electrical connection; the other so-called parasitic elements interact with the electromagnetic field in order to realize a fairly directional antenna but one which is limited to a rather narrow bandwidth. The Yagi antenna has similar looking parasitic dipole elements but which act differently due to their somewhat different lengths. There may be a number of so-called "directors" in front of the active element in the direction of propagation, and usually a single (but possibly more) "reflector" on the opposite side of the active element.

Greater directionality can be obtained using beam-forming techniques such as a parabolic reflector or a horn. Since the size of a directional antenna depends on it being large compared to the wavelength, very directional antennas of this sort are mainly feasible at UHF and microwave frequencies. On the other hand, at low frequencies (such as AM broadcast) where a practical antenna must be much smaller than a wavelength, significant directionality isn't even possible. A vertical antenna or loop antenna small compared to the wavelength is typically used, with the main design challenge being that of impedance matching. With a vertical antenna a *loading coil* at the base of the antenna may be employed to cancel the reactive component of impedance; small loop antennas are tuned with parallel capacitors for this purpose.

An antenna lead-in is the transmission line (or *feed line*) which connects the antenna to a transmitter or receiver. The *antenna feed* may refer to all components connecting the antenna to the transmitter or receiver, such as an impedance matching network in addition to the transmission line. In a so-called aperture antenna, such as a horn or parabolic dish, the "feed" may also refer to a basic antenna inside the entire system (normally at the focus of the parabolic dish or at the throat of a horn) which could be considered the one active element in that antenna system. A microwave antenna may also be fed directly from a waveguide in lieu of a (conductive) transmission line.

An antenna counterpoise or ground plane is a structure of conductive material which improves or substitutes for the ground. It may be connected to or insulated from the natural ground. In a monopole antenna, this aids in the function of the natural ground, particularly where variations (or limitations) of the characteristics of the natural ground interfere with its proper function. Such a structure is normally connected to the return connection of an unbalanced transmission line such as the shield of a coaxial cable.

An electromagnetic wave *refractor* in some aperture antennas is a component which due to its shape and position functions to selectively delay or advance portions of the electromagnetic wavefront passing through it. The refractor alters the spatial characteristics of the wave on one side relative to the other side. It can, for instance, bring the wave to a focus or alter the wave front in other ways, generally in order to maximize the directivity of the antenna system. This is the radio equivalent of an optical lens.

An antenna coupling network is a passive network (generally a combination of inductive and capacitive circuit elements) used for impedance matching in between the antenna and the transmitter or receiver. This may be used to improve the standing wave ratio in order to minimize losses in the transmission line (especially at higher frequencies and/or over longer distances) and to present the transmitter or receiver with a standard resistive impedance (such as 75 ohms) that it expects to see for optimum operation.

Reciprocity

It is a fundamental property of antennas that the characteristics of an antenna described in the next section, such as gain, radiation pattern, impedance, bandwidth, resonant frequency and polarization, are the same whether the antenna is transmitting or receiving. For example, the "*receiving pattern*" (sensitivity as a function of direction) of an antenna when used for reception is identical to the radiation pattern of the antenna when it is *driven* and functions as a radiator. This is a consequence of the reciprocity theorem of electromagnetics. Therefore in discussions of antenna properties no distinction is usually made between receiving and transmitting terminology, and the antenna can be viewed as either transmitting or receiving, whichever is more convenient.

A necessary condition for the above reciprocity property is that the materials in the antenna and transmission medium are linear and reciprocal. *Reciprocal* (or *bilateral*) means that the material has the same response to an electric or magnetic field, or a current, in one direction, as it has to the field or current in the opposite direction. Most materials used in antennas meet these conditions, but some microwave antennas use high-tech components such as isolators and circulators, made of nonreciprocal materials such as ferrite or garnet. These can be used to give the antenna a different behavior on receiving than it has on transmitting, which can be useful in applications like radar.

Parameters

There are several critical parameters affecting an antenna's performance that can be adjusted during the design process. These are resonant frequency, impedance, gain,

aperture or radiation pattern, polarization, efficiency and bandwidth. Transmit antennas may also have a maximum power rating, and receive antennas differ in their noise rejection properties. All of these parameters can be measured through various means.

Resonant frequency

Many types of antenna are tuned to work at one particular frequency, and are effective only over a range of frequencies centered on this frequency, called the resonant frequency. These are called *resonant antennas*. The antenna acts as an electrical resonator. When driven at its resonant frequency, large standing waves of voltage and current are excited in the antenna elements. These large currents and voltages radiate intense electromagnetic waves, so the power radiated by the antenna is maximum at the resonant frequency.

In antennas made of thin linear conductive elements, the length of the driven element(s) determines the resonant frequency. To be resonant, the length of a driven element should typically be either half or a quarter of the wavelength at that frequency; these are called half-wave and quarter-wave antennas. The length referred to is not the physical length, but the electrical length of the element, which is the physical length divided by the velocity factor (the ratio of the speed of wave propagation in the wire to c_0 , the speed of light in a vacuum). Antennas are usually also resonant at multiples (harmonics) of the lowest resonant frequency.

Some antenna designs have multiple resonant frequencies, and some are relatively effective over a very broad range of frequencies. or bandwidth. One commonly known type of wide band antenna is the logarithmic or log-periodic antenna.

The resonant frequency also affects the impedance of the antenna. At resonance, the equivalent circuit of an antenna is a pure resistance, with no reactive component. At frequencies other than the resonant frequencies, the antenna has capacitance or inductance as well as resistance. An antenna can be made resonant at other frequencies besides its natural resonant frequency by compensating for these reactances by adding a loading coil or capacitor in series with it. Other properties of an antenna change with frequency, in particular the radiation pattern, so the antenna's operating frequency may be considerably different from the resonant frequency to optimize other important parameters.

Gain

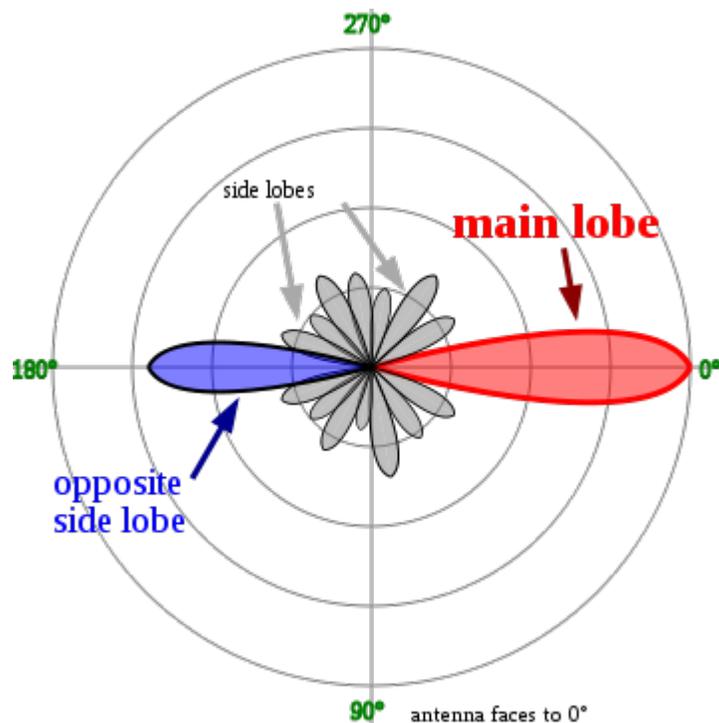
Gain is a parameter which measures the degree of directivity of the antenna's radiation pattern. An antenna with a low gain emits radiation with about the same power in all directions, whereas a high-gain antenna will preferentially radiate in particular directions. Specifically, the *antenna gain*, *directive gain*, or *power gain* of an antenna is defined as the ratio of the intensity (power per unit surface) radiated by the antenna in the direction of its maximum output, at an arbitrary distance, divided by the intensity radiated at the same distance by a hypothetical isotropic antenna.

The gain of an antenna is a passive phenomenon - power is not added by the antenna, but simply redistributed to provide more radiated power in a certain direction than would be transmitted by an isotropic antenna. An antenna designer must take into account the application for the antenna when determining the gain. High-gain antennas have the advantage of longer range and better signal quality, but must be aimed carefully in a particular direction. Low-gain antennas have shorter range, but the orientation of the antenna is relatively inconsequential. For example, a dish antenna on a spacecraft is a high-gain device that must be pointed at the planet to be effective, whereas a typical Wi-Fi antenna in a laptop computer is low-gain, and as long as the base station is within range, the antenna can be in any orientation in space. It makes sense to improve horizontal range at the expense of reception above or below the antenna. Thus most antennas labelled "omnidirectional" really have some gain.

In practice, the half-wave dipole is taken as a reference instead of the isotropic radiator. The gain is then given in **dBd** (decibels over dipole):

NOTE: **0 dBd = 2.15 dBi**. It is vital in expressing gain values that the reference point be included. Failure to do so can lead to confusion and error.

Radiation pattern



polar plots of the horizontal cross sections of a (virtual) Yagi-Uda-antenna. Outline connects points with 3db field power compared to an ISO emitter.

The radiation pattern of an antenna is a plot of the relative field strength of the radio waves emitted by the antenna at different angles. It is typically represented by a three dimensional graph, or polar plots of the horizontal and vertical cross sections. The pattern

of an ideal isotropic antenna, which radiates equally in all directions, would look like a sphere. Many nondirectional antennas, such as monopoles and dipoles, emit equal power in all horizontal directions, with the power dropping off at higher and lower angles; this is called an omnidirectional pattern and when plotted looks like a torus or donut.

The radiation of many antennas shows a pattern of maxima or "*lobes*" at various angles, separated by "*nulls*", angles where the radiation falls to zero. This is because the radio waves emitted by different parts of the antenna typically interfere, causing maxima at angles where the radio waves arrive at distant points in phase, and zero radiation at other angles where the radio waves arrive out of phase. In a directional antenna designed to project radio waves in a particular direction, the lobe in that direction is designed larger than the others and is called the "*main lobe*". The other lobes usually represent unwanted radiation and are called "*sidelobes*". The axis through the main lobe is called the "*principle axis*" or "*boresight axis*".

Impedance

As an electro-magnetic wave travels through the different parts of the antenna system (radio, feed line, antenna, free space) it may encounter differences in impedance (E/H, V/I, etc.). At each interface, depending on the impedance match, some fraction of the wave's energy will reflect back to the source, forming a standing wave in the feed line. The ratio of maximum power to minimum power in the wave can be measured and is called the standing wave ratio (**SWR**). A SWR of 1:1 is ideal. A SWR of 1.5:1 is considered to be marginally acceptable in low power applications where power loss is more critical, although an SWR as high as 6:1 may still be usable with the right equipment. Minimizing impedance differences at each interface (impedance matching) will reduce SWR and maximize power transfer through each part of the antenna system.

Complex impedance of an antenna is related to the electrical length of the antenna at the wavelength in use. The impedance of an antenna can be matched to the feed line and radio by adjusting the impedance of the feed line, using the feed line as an impedance transformer. More commonly, the impedance is adjusted at the load with an antenna tuner, a balun, a matching transformer, matching networks composed of inductors and capacitors, or matching sections such as the gamma match.

Efficiency

Efficiency is the ratio of power actually radiated to the power put into the antenna terminals. A dummy load may have an SWR of 1:1 but an efficiency of 0, as it absorbs all power and radiates heat but very little RF energy, showing that SWR alone is not an effective measure of an antenna's efficiency. Radiation in an antenna is caused by radiation resistance which can only be measured as part of total resistance including loss resistance. Loss resistance usually results in heat generation rather than radiation, and reduces efficiency. Mathematically, efficiency is calculated as radiation resistance divided by total resistance.

Bandwidth

The *bandwidth* of an antenna is the range of frequencies over which it is effective, usually centered on the resonant frequency. The bandwidth of an antenna may be increased by several techniques, including using thicker wires, replacing wires with *cages* to simulate a thicker wire, tapering antenna components (like in a feed horn), and combining multiple antennas into a single assembly (array) and allowing the natural impedance of suitable inductive RF filter traps to select the correct antenna. All these attempts to increase bandwidth by adding capacitance to the surface area have a detrimental effect on efficiency by reducing the Q factor. They also have an adverse effect on the rejection of unwanted harmonics, on both received and transmitted signal frequencies. Small antennas are usually preferred for convenience, but there is a fundamental limit relating bandwidth, size and efficiency.

Polarization

The *polarization* of an antenna is the orientation of the electric field (E-plane) of the radio wave with respect to the Earth's surface and is determined by the physical structure of the antenna and by its orientation. It has nothing in common with antenna directionality terms: "horizontal", "vertical" and "circular". Thus, a simple straight wire antenna will have one polarization when mounted vertically, and a different polarization when mounted horizontally. "Electromagnetic wave polarization filters" are structures which can be employed to act directly on the electromagnetic wave to filter out wave energy of an undesired polarization and to pass wave energy of a desired polarization.

Reflections generally affect polarization. For radio waves the most important reflector is the ionosphere - signals which reflect from it will have their polarization changed unpredictably. For signals which are reflected by the ionosphere, polarization cannot be relied upon. For line-of-sight communications for which polarization can be relied upon, it can make a large difference in signal quality to have the transmitter and receiver using the same polarization; many tens of dB difference are commonly seen and this is more than enough to make the difference between reasonable communication and a broken link.

Polarization is largely predictable from antenna construction but, especially in directional antennas, the polarization of side lobes can be quite different from that of the main propagation lobe. For radio antennas, polarization corresponds to the orientation of the radiating element in an antenna. A vertical omnidirectional WiFi antenna will have vertical polarization (the most common type). An exception is a class of elongated waveguide antennas in which vertically placed antennas are horizontally polarized. Many commercial antennas are marked as to the polarization of their emitted signals.

Polarization is the sum of the E-plane orientations over time projected onto an imaginary plane perpendicular to the direction of motion of the radio wave. In the most general case, polarization is elliptical, meaning that the polarization of the radio waves varies over time. Two special cases are linear polarization (the ellipse collapses into a line) and

circular polarization (in which the two axes of the ellipse are equal). In linear polarization the antenna compels the electric field of the emitted radio wave to a particular orientation. Depending on the orientation of the antenna mounting, the usual linear cases are horizontal and vertical polarization. In circular polarization, the antenna continuously varies the electric field of the radio wave through all possible values of its orientation with regard to the Earth's surface. Circular polarizations, like elliptical ones, are classified as right-hand polarized or left-hand polarized using a "thumb in the direction of the propagation" rule. Optical researchers use the same rule of thumb, but pointing it in the direction of the emitter, not in the direction of propagation, and so are opposite to radio engineers' use.

In practice, regardless of confusing terminology, it is important that linearly polarized antennas be matched, lest the received signal strength be greatly reduced. So horizontal should be used with horizontal and vertical with vertical. Intermediate matchings will lose some signal strength, but not as much as a complete mismatch. Transmitters mounted on vehicles with large motional freedom commonly use circularly polarized antennas so that there will never be a complete mismatch with signals from other sources.

Transmission and reception

All of the antenna parameters are expressed in terms of a transmission antenna, but are identically applicable to a receiving antenna, due to reciprocity. Impedance, however, is not applied in an obvious way; for impedance, the impedance at the load (where the power is consumed) is most critical. For a transmitting antenna, this is the antenna itself. For a receiving antenna, this is at the (radio) receiver rather than at the antenna. Tuning is done by adjusting the length of an electrically long linear antenna to alter the electrical resonance of the antenna.

Antenna tuning is done by adjusting an inductance or capacitance combined with the active antenna (but distinct and separate from the active antenna). The inductance or capacitance provides the reactance which combines with the inherent reactance of the active antenna to establish a resonance in a circuit including the active antenna. The established resonance being at a frequency other than the natural electrical resonant frequency of the active antenna. Adjustment of the inductance or capacitance changes this resonance.

Antennas used for transmission have a maximum power rating, beyond which heating, arcing or sparking may occur in the components, which may cause them to be damaged or destroyed. Raising this maximum power rating usually requires larger and heavier components, which may require larger and heavier supporting structures. This is a concern only for transmitting antennas, as the power received by an antenna rarely exceeds the microwatt range.

Antennas designed specifically for reception might be optimized for noise rejection capabilities. An *antenna shield* is a conductive or low reluctance structure (such as a wire, plate or grid) which is adapted to be placed in the vicinity of an antenna to reduce,

as by dissipation through a resistance or by conduction to ground, undesired electromagnetic radiation, or electric or magnetic fields, which are directed toward the active antenna from an external source or which emanate from the active antenna. Other methods to optimize for noise rejection can be done by selecting a narrow bandwidth so that noise from other frequencies is rejected, or selecting a specific radiation pattern to reject noise from a specific direction, or by selecting a polarization different from the noise polarization, or by selecting an antenna that favors either the electric or magnetic field.

For instance, an antenna to be used for reception of low frequencies (below about ten megahertz) will be subject to both man-made noise from motors and other machinery, and from natural sources such as lightning. Successfully rejecting these forms of noise is an important antenna feature. A small coil of wire with many turns is more able to reject such noise than a vertical antenna. However, the vertical will radiate much more effectively on transmit, where extraneous signals are not a concern.

Basic antenna models



Typical US multiband TV antenna (aerial)

There are many variations of antennas. Below are a few basic models. More can be found in Category:Radio frequency antenna types.

- The isotropic radiator is a purely theoretical antenna that radiates equally in all directions. It is considered to be a point in space with no dimensions and no mass. This antenna cannot physically exist, but is useful as a theoretical model for comparison with all other antennas. Most antennas' gains are measured with reference to an isotropic radiator, and are rated in dBi (decibels with respect to an isotropic radiator).
- The dipole antenna is simply two wires pointed in opposite directions arranged either horizontally or vertically, with one end of each wire connected to the radio and the other end hanging free in space. Since this is the simplest practical antenna, it is also used as a reference model for other antennas; gain with respect to a dipole is labeled as dBd. Generally, the dipole is considered to be omnidirectional in the plane perpendicular to the axis of the antenna, but it has deep nulls in the directions of the axis. Variations of the dipole include the folded dipole, the half wave antenna, the ground plane antenna, the whip, and the J-pole.
- The Yagi-Uda antenna is a directional variation of the dipole with parasitic elements added which are functionality similar to adding a reflector and lenses (directors) to focus a filament light bulb.
- The random wire antenna is simply a very long (at least one quarter wavelength) wire with one end connected to the radio and the other in free space, arranged in any way most convenient for the space available. Folding will reduce effectiveness and make theoretical analysis extremely difficult. (The added length helps more than the folding typically hurts.) Typically, a random wire antenna will also require an antenna tuner, as it might have a random impedance that varies non-linearly with frequency.
- The horn is used where high gain is needed, the wavelength is short (microwave) and space is not an issue. Horns can be narrow band or wide band, depending on their shape. A horn can be built for any frequency, but horns for lower frequencies are typically impractical. Horns are also frequently used as reference antennas.
- The parabolic antenna consists of an active element at the focus of a parabolic reflector to reflect the waves into a plane wave. Like the horn it is used for high gain, microwave applications, such as satellite dishes.
- The patch antenna consists mainly of a square conductor mounted over a groundplane. Another example of a planar antenna is the tapered slot antenna (TSA), as the Vivaldi-antenna.

Practical antennas



"Rabbit ears" set-top antenna

Although any circuit can radiate if driven with a signal of high enough frequency, most practical antennas are specially designed to radiate efficiently at a particular frequency. An example of an inefficient antenna is the simple Hertzian dipole antenna, which radiates over wide range of frequencies and is useful for its small size. A more efficient variation of this is the half-wave dipole, which radiates with high efficiency when the signal wavelength is twice the electrical length of the antenna.

One of the goals of antenna design is to minimize the reactance of the device so that it appears as a resistive load. An "antenna inherent reactance" includes not only the distributed reactance of the active antenna but also the natural reactance due to its location and surroundings (as for example, the capacity relation inherent in the position of the active antenna relative to ground). Reactance diverts energy into the reactive field, which causes unwanted currents that heat the antenna and associated wiring, thereby wasting energy without contributing to the radiated output. Reactance can be eliminated by operating the antenna at its resonant frequency, when its capacitive and inductive reactances are equal and opposite, resulting in a net zero reactive current. If this is not possible, compensating inductors or capacitors can instead be added to the antenna to cancel its reactance as far as the source is concerned.

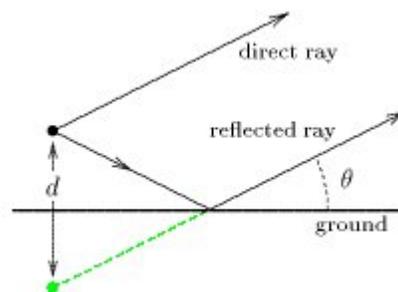
Once the reactance has been eliminated, what remains is a pure resistance, which is the sum of two parts: the ohmic resistance of the conductors, and the radiation resistance. Power absorbed by the ohmic resistance becomes waste heat, and that absorbed by the radiation resistance becomes radiated electromagnetic energy. The greater the ratio of radiation resistance to ohmic resistance, the more efficient the antenna.

Effect of ground

Antennas are typically used in an environment where other objects are present that may have an effect on their performance. Height above ground has a very significant effect on the radiation pattern of some antenna types.

At frequencies used in antennas, the ground behaves mainly as a dielectric. The conductivity of ground at these frequencies is negligible. When an electromagnetic wave arrives at the surface of an object, two waves are created: one enters the dielectric and the other is reflected. If the object is a conductor, the transmitted wave is negligible and the reflected wave has almost the same amplitude as the incident one. When the object is a dielectric, the fraction reflected depends (among others things) on the angle of incidence. When the angle of incidence is small (that is, the wave arrives almost perpendicularly) most of the energy traverses the surface and very little is reflected. When the angle of incidence is near 90° (grazing incidence) almost all the wave is reflected.

Most of the electromagnetic waves emitted by an antenna to the ground below the antenna at moderate (say $< 60^\circ$) angles of incidence enter the earth and are absorbed (lost). But waves emitted to the ground at grazing angles, far from the antenna, are almost totally reflected. At grazing angles, the ground behaves as a mirror. Quality of reflection depends on the nature of the surface. When the irregularities of the surface are smaller than the wavelength reflection is good.



The wave reflected by earth can be considered as emitted by the image antenna

This means that the receptor "sees" the real antenna and, under the ground, the image of the antenna reflected by the ground. If the ground has irregularities, the image will appear fuzzy.

If the receiver is placed at some height above the ground, waves reflected by ground will travel a little longer distance to arrive to the receiver than direct waves. The distance will be the same only if the receiver is close to ground.

In the drawing at right, we have drawn the angle θ far bigger than in reality. Distance between the antenna and its image is d .

The situation is a bit more complex because the reflection of electromagnetic waves depends on the polarization of the incident wave. As the refractive index of the ground (average value $\simeq 2$) is bigger than the refractive index of the air ($\simeq 1$), the direction of the component of the electric field parallel to the ground inverts at the reflection. This is equivalent to a phase shift of π radians or 180° . The vertical component of the electric field reflects without changing direction. This sign inversion of the parallel component and the non-inversion of the perpendicular component would also happen if the ground were a good electrical conductor.



The vertical component of the current reflects without changing sign. The horizontal component reverses sign at reflection.

This means that a receiving antenna "sees" the image antenna with the current in the same direction if the antenna is vertical or with the current inverted if the antenna is horizontal.

For a vertical polarized emission antenna the far electric field of the electromagnetic wave produced by the direct ray plus the reflected ray is:

$$|E_{\perp}| = 2 |E_{\theta_1}| \left| \cos \left(\frac{kd}{2} \sin \theta \right) \right|$$

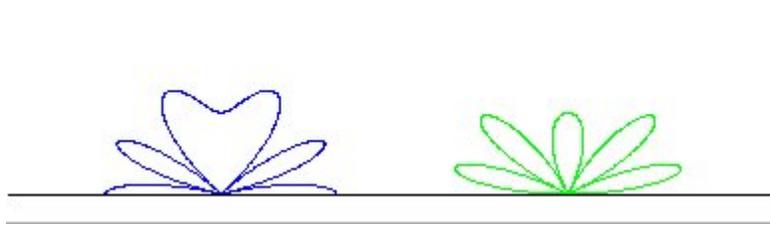
The sign inversion for the parallel field case just changes a cosine to a sine:

$$|E_{\parallel}| = 2 |E_{\theta_1}| \left| \sin \left(\frac{kd}{2} \sin \theta \right) \right|$$

In these two equations:

- E_{θ_1} is the electrical field radiated by the antenna if there were no ground.
- $k = \frac{2\pi}{\lambda}$ is the wave number.
- λ is the wave length.

- d is the distance between antenna and its image (twice the height of the center of the antenna).



Radiation patterns of antennas and their images reflected by the ground. At left the polarization is vertical and there is always a maximum for $\theta=0$. If the polarization is horizontal as at right, there is always a zero for $\theta=0$.

For emitting and receiving antenna situated near the ground (in a building or on a mast) far from each other, distances traveled by direct and reflected rays are nearly the same. There is no induced phase shift. If the emission is polarized vertically the two fields (direct and reflected) add and there is maximum of received signal. If the emission is polarized horizontally the two signals subtracts and the received signal is minimum. This is depicted in the image at right. In the case of vertical polarization, there is always a maximum at earth level (left pattern). For horizontal polarization, there is always a minimum at earth level. Note that in these drawings the ground is considered as a perfect mirror, even for low angles of incidence. In these drawings the distance between the antenna and its image is just a few wavelengths. For greater distances, the number of lobes increases.

Note that the situation is different—and more complex—if reflections in the ionosphere occur. This happens over very long distances (thousands of kilometers). There is not a direct ray but several reflected rays that add with different phase shifts.

This is the reason why almost all public address radio emissions have vertical polarization. As public users are near ground, horizontal polarized emissions would be poorly received. Observe household and automobile radio receivers. They all have vertical antennas or horizontal ferrite antennas for vertical polarized emissions. In cases where the receiving antenna must work in any position, as in mobile phones, the emitter and receivers in base stations use circular polarized electromagnetic waves.

Classical (analog) television emissions are an exception. They are almost always horizontally polarized, because the presence of buildings makes it unlikely that a good emitter antenna image will appear. However, these same buildings reflect the electromagnetic waves and can create ghost images. Using horizontal polarization, reflections are attenuated because of the low reflection of electromagnetic waves whose magnetic field is parallel to the dielectric surface near the Brewster's angle. Vertically polarized analog television has been used in some rural areas. In digital terrestrial television reflections are less obtrusive, due to the inherent robustness of digital signalling and built-in error correction.

- v_i is the voltage applied to the antenna i
- Z_{ii} is the impedance of antenna i
- Z_{ij} is the mutual impedance between antennas i and j

Note that, as is the case for mutual inductances,

$$Z_{ij} = Z_{ji}$$

This is a consequence of Lorentz reciprocity. If some of the elements are not fed (there is a short circuit instead a feeder cable), as is the case in television antennas (Yagi-Uda antennas), the corresponding v_i are zero. Those elements are called parasitic elements. Parasitic elements are unpowered elements that either reflect or absorb and reradiate RF energy.

In some geometrical settings, the mutual impedance between antennas can be zero. This is the case for crossed dipoles used in circular polarization antennas.

Antennas and antenna arrays



A Yagi-Uda beam antenna.



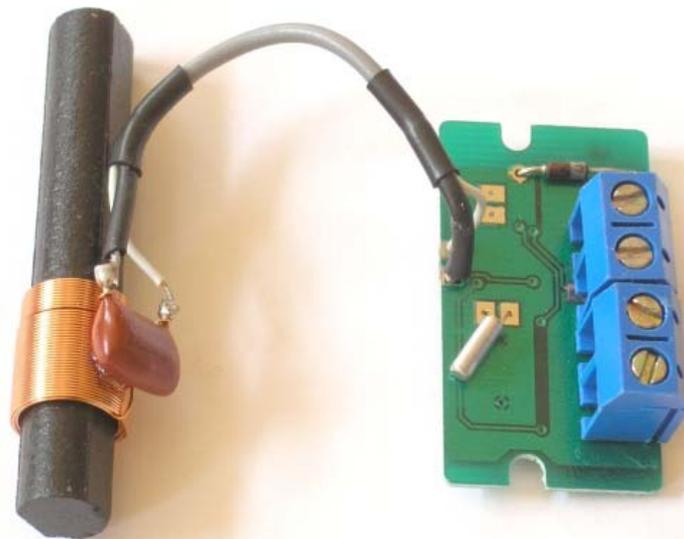
A multi-band rotary directional antenna for amateur radio use.



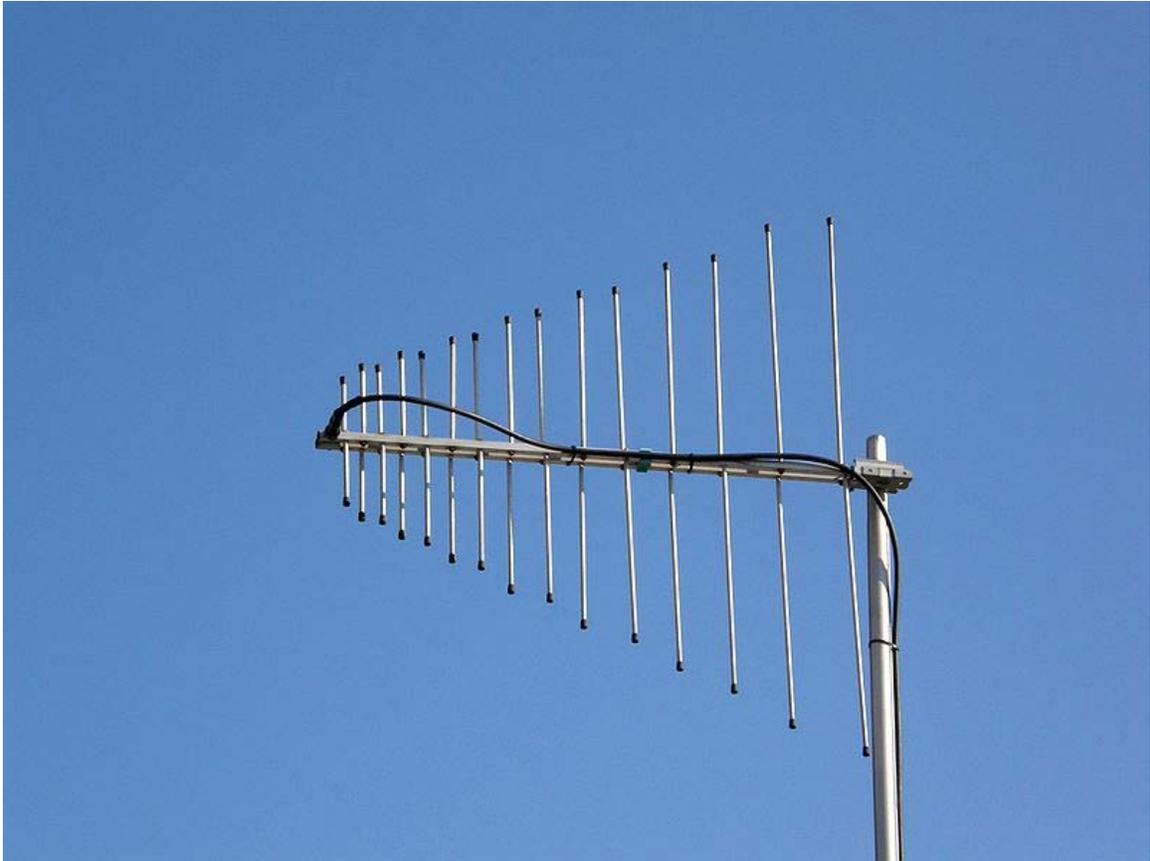
Rooftop TV antenna. It is actually three Yagi antennas. The longest elements are for the low band, while the medium and short elements are for the high and UHF band.



Examples of US 136-174 MHz base station antennas.



Low cost LF time signal receiver, antenna (left) and receiver (right).



Rotatable log-periodic array for VHF and UHF.



Shortwave antennas in Delano, California.



An old VHF-band Yagi-type television antenna.



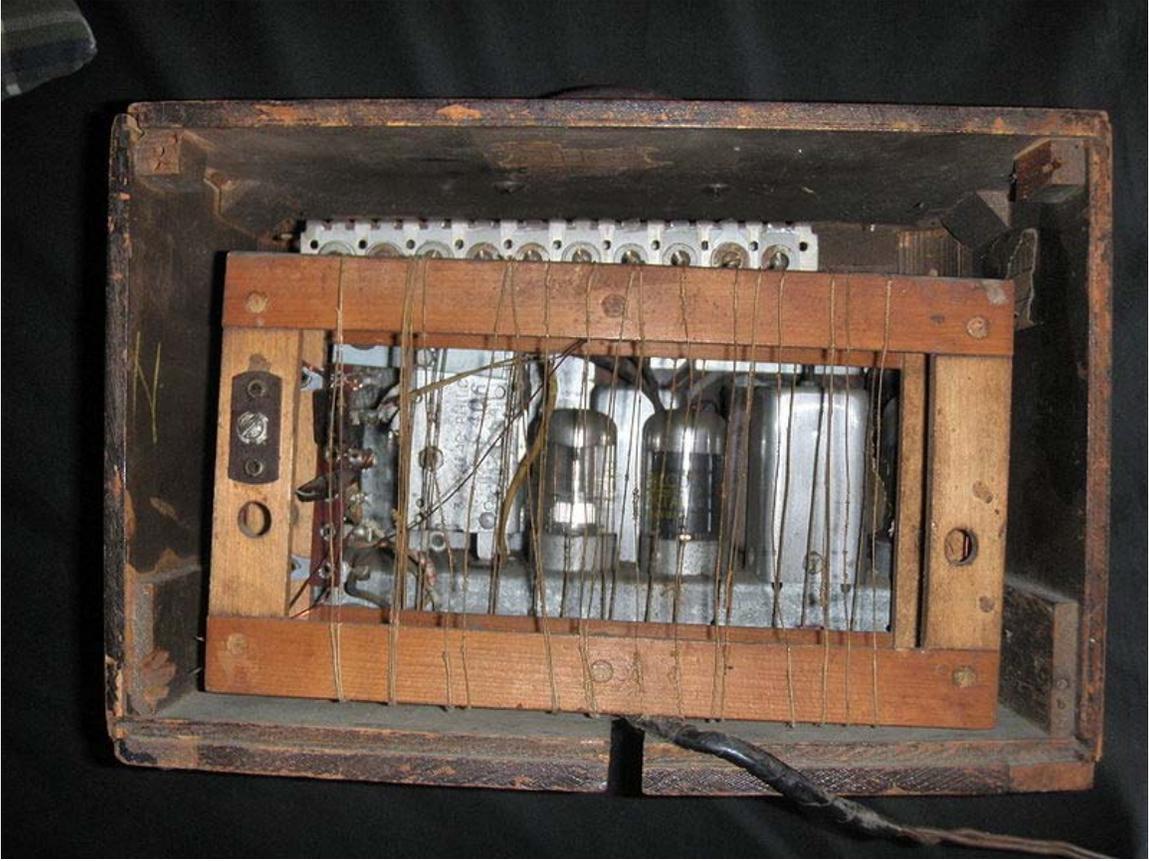
A T2FD broadband antenna, covering the 5-30MHz band.



A US multiband "aerial" TV antenna.



"Rabbit ears" antenna



AM loop antenna

Chapter 8

Receiver (Radio)

A **radio receiver** is an electronic circuit that receives its input from an antenna, uses electronic filters to separate a wanted radio signal from all other signals picked up by this antenna, amplifies it to a level suitable for further processing, and finally converts through demodulation and decoding the signal into a form usable for the consumer, such as sound, pictures, digital data, measurement values, navigational positions, etc.



Early broadcast radio receiver--wireless Truetone model from about 1940

In consumer electronics, the terms *radio* and *radio receiver* are often used specifically for receivers designed for the sound signals transmitted by radio broadcasting services – historically the first mass-market radio application.

Types of radio receivers

Various types of radio receivers may include:

- Consumer audio and high fidelity audio receivers and AV receivers used by home stereo listeners and audio and home theatre system enthusiasts.
- Communications receivers, used as a component of a radio communication link, characterized by high stability and reliability of performance.
- Simple crystal radio receivers (also known as a crystal set) which operate using the power received from radio waves.
- Satellite television receivers, used to receive television programming from communication satellites in geosynchronous orbit.
- Specialized-use receivers such as telemetry receivers that allow the remote measurement and reporting of information.
- Measuring receivers (also: measurement receivers) are calibrated laboratory-grade devices that are used to measure the signal strength of broadcasting stations, the electromagnetic interference radiation emitted by electrical products, as well as to calibrate RF attenuators and signal generators.
- Scanners are specialized receivers that can automatically scan two or more discrete frequencies, stopping when they find a signal on one of them and then continuing to scan other frequencies when the initial transmission ceases. They are mainly used for monitoring VHF and UHF radio systems.
- Internet radio device

Consumer audio receivers

In the context of home audio systems, the term "receiver" often refers to a combination of a tuner, a preamplifier, and a power amplifier all on the same chassis. Audiophiles will refer to such a device as an **integrated receiver**, while a single chassis that implements only one of the three component functions is called a **discrete component**. Some audio purists still prefer three discreet units - tuner, preamplifier and power amplifier - but the integrated receiver has, for some years, been the mainstream choice for music listening. The first integrated stereo receiver was made by the Harman Kardon company, and came onto the market in 1958. It had undistinguished performance, but it represented a breakthrough to the "all in one" concept of a receiver, and rapidly improving designs

gradually made the receiver the mainstay of the marketplace. Many radio receivers also include a loudspeaker.

Hi-Fi / Home theater

Today AV receivers are a common component in a high-fidelity or home-theatre system. The receiver is generally the nerve centre of a sophisticated home-theatre system providing selectable inputs for a number of different audio components like turntables, compact-disc players and recorders, and tape decks (like video-cassette recorders) and video components (DVD players and recorders, video-game systems, and televisions).

With the decline of vinyl discs, modern receivers tend to omit inputs for turntables, which have separate requirements of their own. All other common audio/visual components can use any of the identical line-level inputs on the receiver for playback, regardless of how they are marked (the "name" on each input is mostly for the convenience of the user.) For instance, a second CD player can be plugged into an "Aux" input, and will work the same as it will in the "CD" input jacks.

Some receivers can also provide signal processors to give a more realistic illusion of listening in a concert hall. Digital audio S/PDIF and USB connections are also common today. The home theater receiver, in the vocabulary of consumer electronics, comprises both the 'radio receiver' and other functions, such as control, sound processing, and power amplification. The standalone radio receiver is usually known in consumer electronics as a tuner.

Some modern integrated receivers can send audio out to seven loudspeakers and an additional channel for a subwoofer and often include connections for headphones. Receivers vary greatly in price, and support stereophonic or surround sound. A high-quality receiver for dedicated audio-only listening (two channel stereo) can be relatively inexpensive; excellent ones can be purchased for \$300 United States or less. Because modern receivers are purely electronic devices with no moving parts unlike electromechanical devices like turntables and cassette decks, they tend to offer many years of trouble-free service. In recent years, the home theater in a box has become common, which often integrates a surround-capable receiver with a DVD player. The user simply connects it to a television, perhaps other components, and a set of loudspeakers.

Portable radios

Portable radios include simple transistor radios that are typically monoaural and receive the AM, FM, and/or short wave broadcast bands. FM, and often AM, radios are sometimes included as a feature of portable DVD/CD, MP3 CD, and USB key players, as well as cassette player/recorders.

AM/FM stereo car radios can be a separate dashboard mounted component or a feature of in car entertainment systems.

A Boombox (or Boom-box)—also sometimes known as a GhettoBlaster or a Jambox, or (in parts of Europe) as a "radio-cassette"—is a name given to larger portable stereo systems capable of playing radio stations and recorded music, often at a high level of volume.

Self-powered portable radios, such as clockwork radios are used in developing nations or as part of an emergency preparedness kit.

History of radio receivers

Early development

While James Clerk Maxwell was the first person to prove electromagnetic waves existed, in 1887 a German named Heinrich Hertz demonstrated these new waves by using spark gap equipment to transmit and receive radio or "Hertzian waves", as they were first called. The experiments were not followed up by Hertz. The practical applications of the wireless communication and remote control technology were implemented by Nikola Tesla.

The world's first radio receiver (thunderstorm register) was designed by Alexander Stepanovich Popov, and it was first seen at the All-Russia Exhibition 1896. He was the first to demonstrate the practical application of electromagnetic (radio) waves, although he did not care to apply for a patent for his invention.

A device called a coherer became the basis for receiving radio signals. The first person to use the device to detect radio waves was a Frenchman named Edouard Branly, and Oliver Lodge popularised it when he gave a lecture in 1898 in honour of Hertz. Lodge also made improvements to the coherer. Many experimenters at the time believed that these new waves could be used to communicate over great distances and made significant improvements to both radio receiving and transmitting apparatus. In 1895 Marconi demonstrated the first viable radio system, leading to transatlantic radio communication in December 1901. The honor was later contested as he was found to be using equipment and designs of other experimenters that held the patents at that time.

John Ambrose Fleming's development of an early thermionic valve to help detect radio waves was based upon a discovery of Thomas Edison's (called "The Edison effect", which essentially modified an early light bulb). Fleming called it his "oscillation valve" because it acted in the same way as water valve in only allowing flow in one direction. While Fleming's valve was a great stride forward it would take some years before thermionic, or vacuum tube technology was fully adopted.

Around this time work on other types of detectors started to be undertaken and it resulted in what was later known as the cat's whisker. It consisted of a crystal of a material such as galena with a small springy piece of wire brought up against it. The detector was constructed so that the wire contact could be moved to different points on the crystal, and thereby obtain the best point for rectifying the signal and the best detection. They were

never very reliable as the "whisker" needed to be moved periodically to enable it to detect the signal properly.

Valves (Tubes)

An American named Lee de Forest, a competitor to Marconi, set about to develop receiver technology that did not infringe any patents to which Marconi had access. He took out a number of patents in the period between 1905 and 1907 covering a variety of developments that culminated in the form of the triode valve in which there was a third electrode called a grid. He called this an audion tube. One of the first areas in which valves were used was in the manufacture of telephone repeaters, and although the performance was poor, they gave significant improvement in long distance telephone receiving circuits.

With the discovery that triode valves could amplify signals it was soon noticed that they would also oscillate, a fact that was exploited in generating signals. Once the triode was established as an amplifier it made a tremendous difference to radio receiver performance as it allowed the incoming signals to be amplified. One way that proved very successful was introduced in 1913 and involved the use of positive feedback in the form of a regenerative detector. This gave significant improvements in the levels of gain that could be achieved, greatly increasing selectivity, enabling this type of receiver to outperform all other types of the era. With the outbreak of the First World War, there was a great impetus to develop radio receiving technology further. An American named Irving Langmuir helped introduce a new generation of totally air-evacuated "hard" valves. H. J. Round undertook some work on this and in 1916 he produced a number of valves with the grid connection taken out of the top of the envelope away from the anode connection.

Autodyne and superheterodyne

By the 1920s, the tuned radio frequency receiver (TRF) represented a major improvement in performance over what had been available before, it still fell short of the needs for some of the new applications. To enable receiver technology to meet the needs placed upon it a number of new ideas started to surface. One of these was a new form of direct conversion receiver. Here an internal or local oscillator was used to beat with the incoming signal to produce an audible signal that could be amplified by an audio amplifier.

H. J. Round developed a receiver he called an autodyne in which the same valve was used as a mixer and an oscillator, Whilst the set used fewer valves it was difficult to optimise the circuit for both the mixer and oscillator functions.

The next leap forward in receiver technology was a new type of receiver known as the superheterodyne, or supersonic heterodyne receiver. A Frenchman named Lucien Levy was investigating ways in which receiver selectivity could be improved and in doing this he devised a system whereby the signals were converted down to a lower frequency where the filter bandwidths could be made narrower. A further advantage was that the

gain of valves was considerably greater at the lower frequencies used after the frequency conversion, and there were fewer problems with the circuits bursting into oscillation.

The idea for developing a receiver with a fixed intermediate frequency amplifier and filter is credited to Edwin Armstrong of the United States. Working for the American Expeditionary Force in Europe during 1918, Armstrong thought that if the incoming signals was mixed with a variable frequency oscillator (the "local oscillator"), a lower-frequency fixed tuned amplifier could be used. Armstrong's original receiver consisted of a total of eight vacuum tubes. Several tuned circuits could be cascaded to improve selectivity, and being set on a fixed frequency they did not all need to be changed in line with one another. The filters could be preset and left correctly tuned. Armstrong was not the only person working on the idea of a superheterodyne receiver. Alexander Meissner in Germany had taken out a patent for the idea six months before Armstrong, but since Meissner did not prove the idea in practice, and he did not build a superheterodyne radio, the invention is credited to Armstrong.

The need for the increased performance of the superheterodyne receiver was first experienced in North America, and by the late 1920s most radio sets were superheterodyne receivers. However, in Europe the number of broadcast stations did not start to rise as rapidly until later. Even so, by the mid-1930s virtually all receiving sets in Europe as well were using the superheterodyne principle. In 1926, the tetrode valve was introduced, and enabled further improvements in performance.

War and postwar developments



Military HF receiver, type BC-224-D (1942)

In 1939 the outbreak of war gave a new impetus to receiver development. During this time a number of classic communications receivers were designed. Some like the National HRO are still sought by enthusiasts today and although they are relatively large by today's standards, they can still give a good account of themselves under current crowded band conditions. In the late 1940s the transistor was discovered. Initially the devices were not widely used because of their expense, and the fact that valves were being made smaller, and performed better. However by the early 1960s portable transistor broadcast receivers (transistor radios) were hitting the market place. These radios were ideal for broadcast reception on the long and medium wave bands. They were much smaller than their valve equivalents, they were portable and could be powered from batteries. Although some valve portable receivers were available, batteries for these were expensive and did not last for long. The power requirements for transistor radios were very much less, resulting in batteries lasting for much longer and being considerably cheaper.

Semiconductors

Further developments in semiconductor technology led to the introduction of the integrated circuit in the late 1950s. This enabled radio receiver technology to move forward even further. Integrated circuits enabled high performance circuits to be built for less cost, and significant amounts of space could be saved.

As a result of these developments new techniques could be introduced. One of these was the frequency synthesizer that was used to generate the local oscillator signal for the receiver. By using a synthesizer it was possible to generate a very accurate and stable local oscillator signal. Also the ability of synthesizers to be controlled by microprocessors meant that many new facilities could be introduced apart from the significant performance improvements offered by synthesizers.

Digital technologies

Receiver technology had been advancing gradually and regularly. Many of the functions performed by analogue electronics can be performed by software instead. The benefit is that software is not affected by temperature, physical variables, electronic noise and manufacturing defects.

While today's radios are amazing pieces of modern technology, filled with low- power, high performance, integrated circuits crammed into the smallest spaces, the basic principle of the radio receiver is practically always the superheterodyne one, the same idea which was developed by Edwin Armstrong back in 1918. For really high-performance receivers, such as satellite communications receivers and military/naval receivers, two-stage ("double conversion") and even three-stage ("triple conversion") superheterodyne processing is frequently used. Single-conversion receivers are rather simple-minded in their nature.

DSP technology

DSP technology, short for digital signal processing, is the use of digital means to process signals and is coming into wide use in modern shortwave receivers. It is the basis of many areas of modern technology including cell phones, CD players, video recorders and computers. A digital signal is essentially a stream or sequence of numbers that relay a message through some sort of medium such as a wire. The primary benefit of DSP hardware in shortwave receivers is the ability to tailor the bandwidth of the receiver to current reception conditions and to the type of signal being listened to. A typical analog only receiver may have a limited number of fixed bandwidths, or only one, but a DSP receiver may have 40 or more individually selectable filters.

PC controlled radio receivers

"PC radios", or radios that are designed to be controlled by a standard PC are controlled by specialized PC software using a serial port connected to the radio. A "PC radio" may not have a front-panel at all, and may be designed exclusively for computer control, which reduces cost.

Some PC radios have the great advantage of being field upgradable by the owner. New versions of the DSP firmware can be downloaded from the manufacturer's web site and uploaded into the flash memory of the radio. The manufacturer can then in effect add new features to the radio over time, such as adding new filters, DSP noise reduction, or simply to correct bugs.

A full-featured radio control program allows for scanning and a host of other functions and, in particular, integration of databases in real-time, like a "TV-Guide" type capability. This is particularly helpful in locating all transmissions on all frequencies of a particular broadcaster, at any given time. Some control software designers have even integrated Google Earth to the shortwave databases, so it is possible to "fly" to a given transmitter site location with a click of a mouse. In many cases the user is able to see the transmitting antennas where the signal is originating from.

Radio control software

The field of software control of PC radios has grown rapidly in the last several years, with developers making a number of advances. Since the Graphical User Interface or GUI interface PC to the radio has unlimited flexibility, any number of new features can be added by the software designer. Features that can be found in advanced control software programs today include a band table, GUI controls corresponding to traditional radio controls, local time clock and a UTC clock, signal strength meter, an ILG database for shortwave listening with lookup capability, scanning capability, text-to-speech interface, and integrated Conference Server.

Software-defined radios

The next level in radio / software integration are so-called pure "software defined radios". The distinction here is that all filtering, modulation and signal manipulation is done in software, usually by a PC soundcard or by a dedicated piece of DSP hardware. There may be a minimal RF front-end or traditional radio that supplies an IF to the SDR. SDR's can go far beyond the usual demodulation capability of typical, and even high-end DSP shortwave radios. They can for example, record large swaths of the radio spectrum to a hard drive for "playback" at a later date. The same SDR that one minute is demodulating a simple AM broadcast may also be able to decode an HDTV broadcast in the next. A well known open-source project called GNU Radio is dedicated to evolving a high-performance SDR. All the source code for this SDR is freely downloadable and modifiable by anyone.

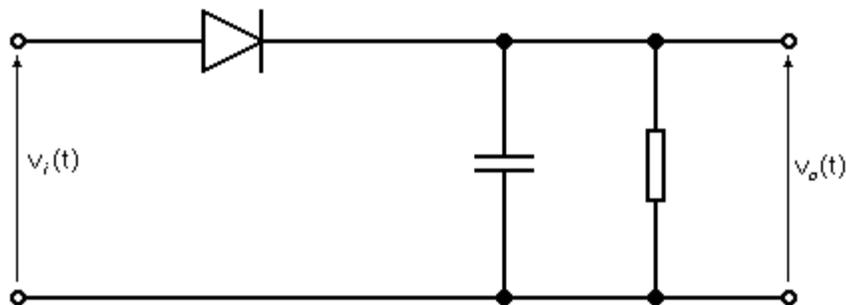
Chapter 9

Detector

A **detector** is a device that recovers information of interest contained in a modulated wave. The term dates from the early days of radio when all transmissions were in Morse code, and it was only necessary to *detect* the presence (or absence) of a radio wave using a device such as a coherer without necessarily making it audible. A more up-to-date term is "demodulator", but "detector" has a history of many decades of use, even if it is a misnomer.

Amplitude modulation detectors

Envelope detector



A simple envelope detector

One major technique is known as envelope detection. The simplest form of envelope detector is the diode detector that consists of a diode connected between the input and output of the circuit, with a resistor and capacitor in parallel from the output of the circuit to the ground. If the resistor and capacitor are correctly chosen, the output of this circuit will approximate a voltage-shifted version of the original signal.

An early form of envelope detector was the cat's whisker, which was used in the crystal set radio receiver.

Product detector

A product detector is a type of demodulator used for AM and SSB signals. Rather than converting the envelope of the signal into the decoded waveform like an envelope detector, the product detector takes the product of the modulated signal and a local oscillator, hence the name. At least partially, it multiplies the signal by the output of the local oscillator. This can be accomplished by heterodyning. The received signal is mixed, in some type of nonlinear device, with a signal from the local oscillator, to produce an intermediate frequency, referred to as the beat frequency, from which the modulating signal is detected and recovered.

Frequency and phase modulation detectors

AM detectors cannot demodulate FM and PM signals because both have a constant amplitude. However an AM radio may detect the sound of an FM broadcast by the phenomenon of **slope detection** which occurs when the radio is tuned slightly above or below the nominal broadcast frequency. Frequency variation on one sloping side of the radio tuning curve gives the amplified signal a corresponding local amplitude variation, to which the AM detector is sensitive. Slope detection gives inferior distortion and noise rejection compared to the following dedicated FM detectors that are normally used.

Phase detector

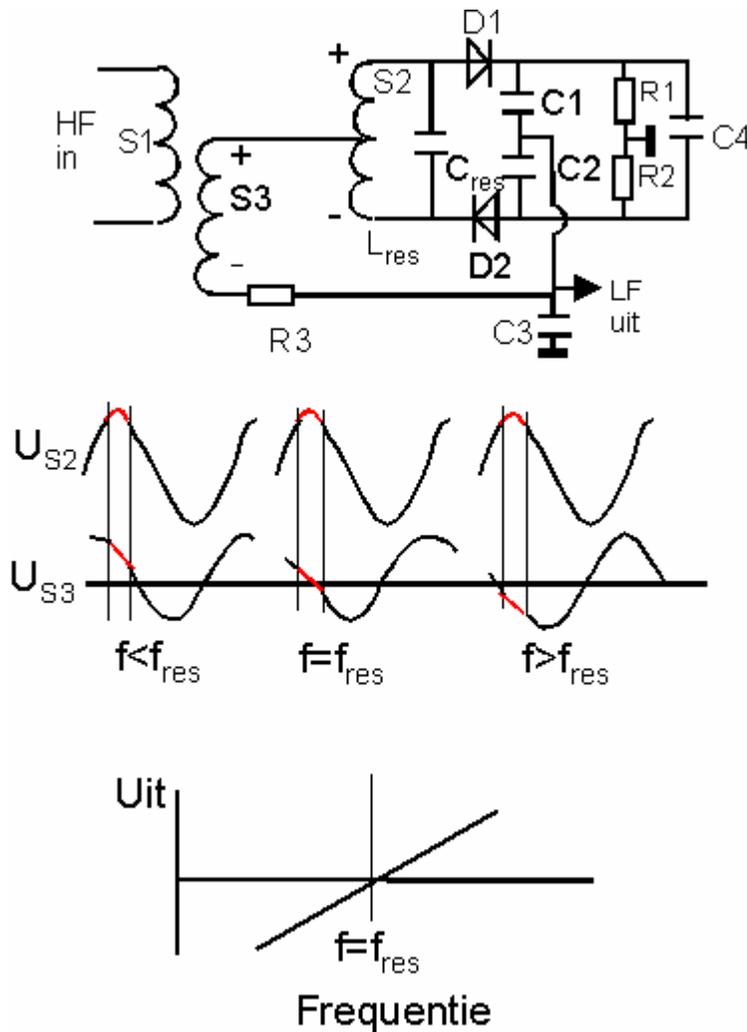
A phase detector is a nonlinear device whose output represents the phase difference between the two oscillating input signals. It has two inputs and one output: a reference signal is applied to one input and the phase or frequency modulated signal is applied to the other. The output is a signal that is proportional to the phase difference between the two inputs.

In phase demodulation the information is contained in the amount and rate of phase shift in the carrier wave.

The Foster-Seeley discriminator

The Foster-Seeley discriminator is a widely used FM detector. The detector consists of a special center-tapped transformer feeding two diodes in a full wave DC rectifier circuit. When the input transformer is tuned to the signal frequency, the output of the discriminator is zero. When there is no deviation of the carrier, both halves of the center tapped transformer are balanced. As the FM signal swings in frequency above and below the carrier frequency, the balance between the two halves of the center-tapped secondary is destroyed and there is an output voltage proportional to the frequency deviation.

Ratio detector



A ratio detector using solid-state diodes

The ratio detector is a variant of the Foster-Seeley discriminator, but one diode conducts in an opposite direction. The output in this case is taken between the sum of the diode voltages and the center tap. The output across the diodes is connected to a large value capacitor, which eliminates AM noise in the ratio detector output. While distinct from the Foster-Seeley discriminator, the ratio detector will similarly not respond to AM signals, however the output is only 50% of the output of a discriminator for the same input signal.

Quadrature detector

In quadrature detectors, the received FM signal is split into two signals. One of the two signals is then passed through a high-reactance capacitor, which shifts the phase of that signal by 90 degrees. This phase-shifted signal is then applied to an LC circuit, which is resonant at the FM signal's unmodulated, "center," or "carrier" frequency. If the received FM signal's frequency equals the center frequency, then the two signals will have a 90-

degree phase difference and they are said to be in "phase quadrature" — hence the name of this method. The two signals are then multiplied together in an analog or digital device, which serves as a phase detector; that is, a device whose output is proportional to the phase difference between two signals. In the case of an unmodulated FM signal, the phase detector's output is — after the output has been filtered; that is, averaged over time — constant; namely, zero. However, if the received FM signal has been modulated, then its frequency will vary from the center frequency. In this case, the resonant LC circuit will further shift the phase of the signal from the capacitor, so that the signal's total phase shift will be the sum of the 90 degrees that's imposed by the capacitor and the positive or negative phase change that's imposed by the LC circuit. Now the output from the phase detector will differ from zero, and in this way, one recovers the original signal that was used to modulate the FM carrier.

This detection process can also be accomplished by combining, in an exclusive-OR (XOR) logic gate, the original FM signal and a square wave whose frequency equals the FM signal's center frequency. The XOR gate produces an output pulse whose duration equals the difference between the times at which the square wave and the received FM signal pass through zero volts. As the FM signal's frequency varies from its unmodulated center frequency (which is also the frequency of the square wave), the output pulses from the XOR gate become longer or shorter. (In essence, this quadrature detector converts an FM signal into a pulse-width modulated (PWM) signal.) When these pulses are filtered, the filter's output rises as the pulses grow longer and its output falls as the pulses grow shorter. In this way, one recovers the original signal that was used to modulate the FM carrier.

Other FM detectors

Less common, specialized, or obsolescent types of detectors include :

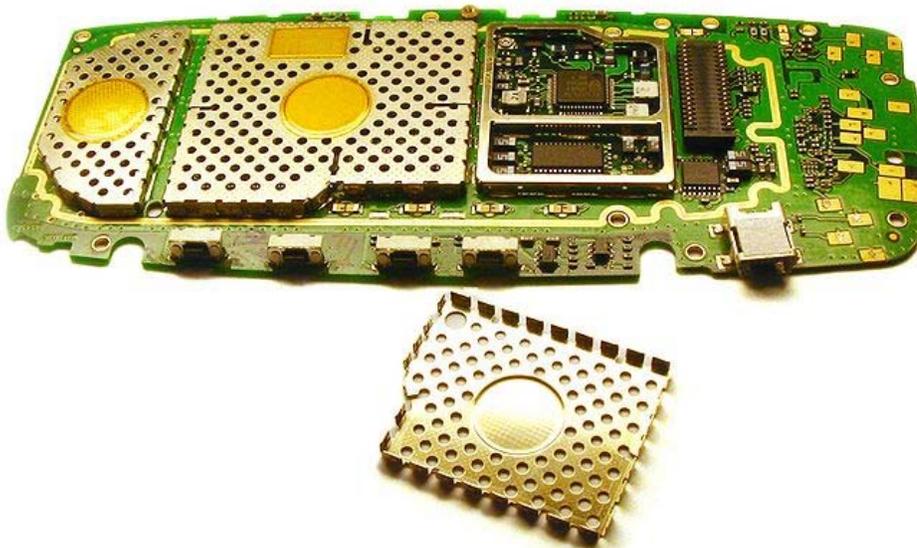
- Travis or double tuned circuit discriminator using two non-interacting tuned circuits above and below the nominal center frequency
- Weiss discriminator which uses a single LC tuned circuit or crystal
- Pulse count discriminator which converts the frequency to a train of constant amplitude pulses, producing a voltage directly proportional to the frequency.

Phase-locked loop detector

The phase-locked loop detector requires no frequency-selective LC network to accomplish demodulation. In this system, a voltage controlled oscillator (VCO) is phase locked by a feedback loop, which forces the VCO to follow the frequency variations of the incoming FM signal. The low-frequency error voltage that forces the VCO's frequency to track the frequency of the modulated FM signal is the demodulated audio output.

Chapter 10

Electromagnetic Shielding



Electromagnetic shielding cages inside a disassembled mobile phone.

Electromagnetic shielding is the process of reducing the electromagnetic field in a space by blocking the field with barriers made of conductive and/or magnetic materials. Shielding is typically applied (1) to enclosures to isolate electrical devices from the 'outside world' and (2) to cables to isolate wires from the environment through which the cable runs. Electromagnetic shielding that blocks radio frequency electromagnetic radiation is also known as **RF shielding**.

The shielding can reduce the coupling of radio waves, electromagnetic fields and electrostatic fields, though not static or low-frequency magnetic fields (a conductive

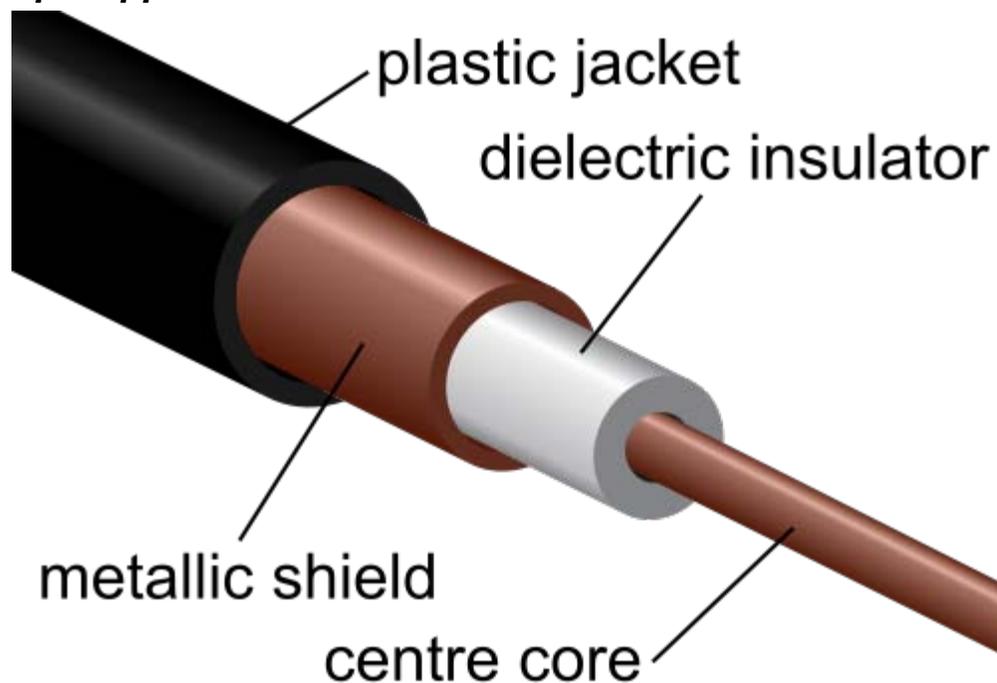
enclosure used to block electrostatic fields is also known as a Faraday cage). The amount of reduction depends very much upon the material used, its thickness, the size of the shielded volume and the frequency of the fields of interest and the size, shape and orientation of apertures in a shield to an incident electromagnetic field.

Materials used

Typical materials used for electromagnetic shielding include sheet metal, metal screen, and metal foam. Any holes in the shield or mesh must be significantly smaller than the wavelength of the radiation that is being kept out, or the enclosure will not effectively approximate an unbroken conducting surface.

Another commonly used shielding method, especially with electronic goods housed in plastic enclosures, is to coat the inside of the enclosure with a metallic ink or similar material. The ink consists of a carrier material loaded with a suitable metal, typically copper or nickel, in the form of very small particulates. It is sprayed on to the enclosure and, once dry, produces a continuous conductive layer of metal, which can be electrically connected to the chassis ground of the equipment, thus providing effective shielding.

Example applications



Cross-section through a coaxial cable showing shielding and other layers

One example is a shielded cable, which has electromagnetic shielding in the form of a wire mesh surrounding an inner core conductor. The shielding impedes the escape of any signal from the core conductor, and also signals from being added to the core conductor.

Some cables have two separate coaxial screens, one connected at both ends, the other at one end only, to maximize shielding of both electromagnetic and electrostatic fields.

The door of a microwave oven has a screen built into the window. From the perspective of microwaves (with wavelengths of 12 cm) this screen finishes a Faraday cage formed by the oven's metal housing. Visible light, with wavelengths ranging between 400 nm and 700 nm, passes easily between the wires.

RF shielding is also used to prevent access to data stored on RFID chips embedded in various devices, such as biometric passports.

NATO specifies electromagnetic shielding for computers and keyboards to prevent passive monitoring of keyboard emissions that would allow passwords to be captured; consumer keyboards do not offer this protection primarily because of the prohibitive cost.

RF shielding is also used to protect medical and laboratory equipment to provide protection against interfering signals, including AM, FM, TV, emergency services, dispatch, pagers, ESMR, cellular, and PCS. It can also be used to protect the equipment at the AM, FM or TV broadcast facilities.

How electromagnetic shielding works

Electromagnetic radiation consists of coupled electric and magnetic fields. The electric field produces forces on the charge carriers (i.e., electrons) within the conductor. As soon as an electric field is applied to the surface of an ideal conductor, it induces a current that causes displacement of charge inside the conductor that cancels the applied field inside, at which point the current stops.

Similarly, *varying* magnetic fields generate eddy currents that act to cancel the applied magnetic field. (The conductor does not respond to static magnetic fields unless the conductor is moving relative to the magnetic field.) The result is that electromagnetic radiation is reflected from the surface of the conductor: internal fields stay inside, and external fields stay outside.

Several factors serve to limit the shielding capability of real RF shields. One is that, due to the electrical resistance of the conductor, the excited field does not completely cancel the incident field. Also, most conductors exhibit a ferromagnetic response to low-frequency magnetic fields, so that such fields are not fully attenuated by the conductor. Any holes in the shield force current to flow around them, so that fields passing through the holes do not excite opposing electromagnetic fields. These effects reduce the field-reflecting capability of the shield.

In the case of high-frequency electromagnetic radiation, the above-mentioned adjustments take a non-negligible amount of time. But then the radiation energy, as far as it is not reflected, is absorbed by the skin (unless it is extremely thin), so in this case there

is no electromagnetic field inside either. This is called the skin effect. A measure for the depth to which radiation can penetrate the shield is the so-called skin depth.

Magnetic shielding

Equipment sometimes requires isolation from external magnetic fields. For static or slowly varying magnetic fields (below about 100 kHz) the Faraday shielding described above is ineffective. There exists a limited possibility of passively isolating a volume magnetically by using shields made of high magnetic permeability metal alloys such as large crystalline grain structure foils or sheet metals of Permalloy and Mu-metal, or with nanocrystalline grain structure ferromagnetic metal coatings. These materials don't block the magnetic field, as with electric shielding, but rather draw the field into themselves, providing a path for the magnetic field lines around the shielded volume. The best shape for magnetic shields is thus a closed container. The effectiveness of this type of shielding decreases with the material's permeability, which generally drops off at both very low magnetic field strengths, and also at high field strengths where the material becomes saturated. So to achieve low residual fields, magnetic shields often consist of several enclosures one inside the other, each of which successively reduces the field inside it.

Because of the above limitations of passive shielding, an alternative used with static fields is active shielding; using a field created by another magnet to cancel out the ambient field within a volume. Solenoids designed to do this are called Helmholtz coils.

Chapter 11

Radio-Frequency Identification

Radio-frequency identification (RFID) is a technology that uses communication via radio waves to exchange data between a reader and an electronic tag attached to an object, for the purpose of identification and tracking.

It is possible in the near future, RFID technology will continue to proliferate in our daily lives the way that bar code technology did over the forty years leading up to the turn of the 21st century bringing unobtrusive but remarkable changes when it was new.

RFID makes it possible to give each product in a grocery store its own unique identifying number, to provide assets, people, work in process, medical devices etc. all with individual unique identifiers - like the license plate on a car but for every item in the world. This is a vast improvement over paper and pencil tracking or bar code tracking that has been used since the 1970s. With bar codes, it is only possible to identify the brand and type of package in a grocery store, for instance. Furthermore, passive RFID tags (those without a battery) can be read if passed within close enough proximity to an RFID reader. It is not necessary to "show" them to it, as with a bar code. In other words it does not require line of sight to "see" an RFID tag, the tag can be read inside a case, carton, box or other container, and unlike barcodes RFID tags can be read hundreds at a time. Bar codes can only read one at a time.

Some RFID tags can be read from several meters away and beyond the line of sight of the reader. The application of bulk reading enables an almost-parallel reading of tags.

Radio-frequency identification involves the hardware known as *interrogators* (also known as *readers*), and *tags* (also known as *labels*), as well as RFID software or RFID middleware.

Most RFID tags contain at least two parts: one is an integrated circuit for storing and processing information, modulating and demodulating a radio-frequency (RF) signal, and other specialized functions; the other is an antenna for receiving and transmitting the signal.

RFID can be either passive (using no battery), active (with an on-board battery that always broadcasts or beacons its signal) or battery assisted passive "BAP" which has a small battery on board that is activated when in the presence of an RFID reader. Passive tags in 2011 start at \$.05 each and for special tags meant to be mounted on metal, or withstand gamma sterilization go up to \$5. Active tags for tracking containers, medical assets, or monitoring environmental conditions in data centers all start at \$50 and can go up over \$100 each. BAP tags are in the \$3-10 range and also have sensor capability like temperature and humidity.

The term RFID refers to the technology. The tags should properly be called "RFID tags" not "RFIDs".

Fixed RFID and Mobile RFID: Depending on mobility, RFID readers are classified into two different types: **fixed RFID and mobile RFID**. If the reader reads tags in a stationary position, it is called fixed RFID. These fixed readers are set up specific interrogation zones and create a "bubble" of RF energy that can be tightly controlled if the physics is well engineered. This allows a very definitive reading area for when tags go in and out of the interrogation zone. On the other hand, if the reader is mobile when the reader reads tags, it is called mobile RFID. Mobile readers include hand helds, carts and vehicle mounted RFID readers from manufacturers such as Motorola, Intermec, Impinj, Sirit, etc.

There are three types of RFID tags: passive RFID tags, which have no power source and require an external electromagnetic field to initiate a signal transmission, active RFID tags, which contain a battery and can transmit signals once an external source ('Interrogator') has been successfully identified, and battery assisted passive (BAP) RFID tags, which require an external source to wake up but have significant higher forward link capability providing greater range.

There are a variety of groups defining standards and regulating the use of RFID, including the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC), ASTM International, the DASH7 Alliance and EPCglobal. (Refer to Regulation and standardization below.) There are also several specific industries that have set guidelines including the Financial Services Technology Consortium (FSTC) has set a standard for tracking IT Assets with RFID, the Computer Technology Industry Association CompTIA has set a standard for certifying RFID engineers and the International Airlines Transport Association IATA set tagging guidelines for luggage in airports.

RFID has many applications; for example, it is used in enterprise supply chain management to improve the efficiency of inventory tracking and management. The Healthcare industry has used RFID to create tremendous productivity increases by eliminating "parasitic" roles that don't add value to an organization such as counting, looking for things, or auditing items. Many financial institutions use RFID to track key assets and automate Sarbanes Oxley SOX compliance. Also with recent advances in

social media RFID is being used to tie the physical world with the virtual world. RFID in Social Media first came to light in 2010 with Facebook's annual conference (f8).

History and technology background



An RFID tag used for electronic toll collection.

In 1945 Léon Theremin invented an espionage tool for the Soviet Union which retransmitted incident radio waves with audio information. Sound waves vibrated a diaphragm which slightly altered the shape of the resonator, which modulated the reflected radio frequency. Even though this device was a covert listening device, not an identification tag, it is considered to be a predecessor of RFID technology, because it was likewise passive, being energized and activated by electromagnetic waves from an outside source.

Similar technology, such as the IFF transponder developed in the United Kingdom, was routinely used by the allies in World War II to identify aircraft as friend or foe. Transponders are still used by most powered aircraft to this day. Another early work exploring RFID is the landmark 1948 paper by Harry Stockman, titled "Communication by Means of Reflected Power" (Proceedings of the IRE, pp 1196–1204, October 1948). Stockman predicted that "... considerable research and development work has to be done before the remaining basic problems in reflected-power communication are solved, and before the field of useful applications is explored."

Mario Cardullo's device in 1973 was the first true ancestor of modern RFID, as it was a passive radio transponder with memory. The initial device was passive, powered by the interrogating signal, and was demonstrated in 1971 to the New York Port Authority and other potential users and consisted of a transponder with 16 bit memory for use as a toll device. The basic Cardullo patent covers the use of RF, sound and light as transmission media. The original business plan presented to investors in 1969 showed uses in transportation (automotive vehicle identification, automatic toll system, electronic license plate, electronic manifest, vehicle routing, vehicle performance monitoring), banking (electronic check book, electronic credit card), security (personnel identification, automatic gates, surveillance) and medical (identification, patient history).

An early demonstration of *reflected power* (modulated backscatter) RFID tags, both passive and semi-passive, was performed by Steven Depp, Alfred Koelle, and Robert

Freyman at the Los Alamos National Laboratory in 1973. The portable system operated at 915 MHz and used 12-bit tags. This technique is used by the majority of today's UHFID and microwave RFID tags.

The first patent to be associated with the abbreviation RFID was granted to Charles Walton in 1983.

The largest deployment of active RFID is the US Department of Defense use of Savi active tags on every one of its more than a million shipping containers that travel outside of the continental United States. The largest passive RFID deployment is the enterprise-wide deployment performed by Wal*Mart which instrumented over 2800 retail stores with over 25,000 reader systems, however the exact number is considered 'corporate confidential'.

Miniaturization

RFIDs are easy to conceal or incorporate in other items. For example, in 2009 researchers at Bristol University successfully glued RFID micro-transponders to live ants in order to study their behavior. This trend towards increasingly miniaturized RFIDs is likely to continue as technology advances.

Hitachi holds the record for the smallest RFID chip, at 0.05mm x 0.05mm. This is 1/64th the size of the previous record holder, the mu-chip. Manufacture is enabled by using the silicon-on-insulator (SOI) process. These dust-sized chips can store 38-digit numbers using 128-bit Read Only Memory (ROM). A major challenge is the attachment of the antennas, thus limiting read range to only millimeters.

Potential alternatives to the radio frequencies (0.125–0.1342, 0.140–0.1485, 13.56, and 840–960 MHz) used are seen in optical RFID (or OPID) at 333 THz (900 nm), 380 THz (788 nm), 750 THz (400 nm). The awkward antennas of RFID can be replaced with photovoltaic components and IR-LEDs on the ICs.

Current uses

In 2010 three key factors drove a significant increase in RFID usage: decreased cost of equipment and tags, increased performance to a reliability of 99.9% and a stable international standard around UHF passive. At RFID Journal Live 2010 in Orlando, Airbus detailed 16 active projects, IBM and — most recently added to the team — CSC. The two other areas of significant use are financial services for IT asset tracking and healthcare with more than 60% of the top medical device companies using passive UHF RFID in 2010. RFID is becoming increasingly prevalent as the price of the technology decreases. The Japanese HIBIKI initiative aims to reduce the price to 5 Yen (4 eurocents). In January 2009 Envego announced a 5.9 cent tag, and in March 2010 a Korean laboratory successfully created a printed chip using carbon nanotubes that would halve the price of a passive UHF RFID tag to about three cents by late 2011.

Payment by mobile phones

Since summer 2009, two credit card companies have been working with Dallas, Texas-based DeviceFidelity to develop specialized microSD cards. When inserted into a mobile phone, the microSD card can be both a passive tag and an RFID reader. After inserting the microSD, a user's phone can be linked to bank accounts and used in mobile payment.

Dairy Queen in conjunction with Vivotech has also begun using RFIDs on mobile phones as part of their new loyalty and rewards program. Patrons can ask to receive an RFID tag to place on their phone. After activation, the phone can receive promotions and coupons, which can be read by ViVOtech's specialized NFC devices.

Similarly, 7-Eleven has been working alongside MasterCard to promote a new touch-free payment system. Those joining the trial are given a complimentary Nokia 3220 cell phone – after activation, it can be used as an RFID-capable MasterCard credit card at any of 7-Eleven's worldwide chains.

Nokia's 2008 device, the 6212, has RFID capabilities also. Credit card information can be stored, and bank accounts can be directly accessed using the enabled handset. The phone, if used as a vector for mobile payment, has added security in that users would be required to enter a passcode or PIN before payment is authorized.

Transportation payments

Governments use RFID applications for traffic management, while automotive companies use various RFID tracking solutions for product management. Many of these solutions may work together in the future, though privacy regulations prevent many initiatives from moving forward at the same pace that technology allows.

Car-sharing

The Zipcar car-sharing service uses RFID cards for locking and unlocking cars and for member identification.

Season parking tickets

Following a successful pilot, Housing & Development Board (HDB) Singapore called two tenders in 2006 to implement RFID to replace the paper Season Parking Ticket (SPT). The successful tenders have distributed RFID tags to SPT holders since March 2007.

Toll roads

- In Pakistan, RFID is being used for e-tolling in Motorways, implemented by NADRA.

- In Dubai, UAE, RFID is being used for e-tolling – SALIK in Motorways, implemented by RTA.
- In Turkey, 13.56 MHz RFID cards have been used in the motorways and bridges as a payment system since 2005; it is also used in public transportation systems in Istanbul, Çanakkale, Izmir and Denizli.
- RFID is used in Malaysia's Touch 'n Go. As the system's name indicates, the card is designed to only function as an RFID card when the user touches it.
- In Norway, all public toll roads are equipped with an RFID payment system known as AutoPASS.
- In Italy, all public toll roads are equipped with an optional RFID payment system named Telepass.
- In Greece, public toll roads in the Attica and Peloponnese regions are equipped with an RFID payment system known as E-Pass.
- In Ireland, the eToll system uses RFID tags for payments on all road tolls, including the barrier-free M50 toll between exits 6 and 7.
- In Singapore, public transportation buses and trains employ passive RFID cards known as EZ-Link cards. Traffic into crowded downtown areas is regulated by variable tolls imposed using an active tagging system combined with the use of stored-value cards (known as CashCards).
- In Ontario, Canada, Electronic Road Pricing systems are used to collect toll payments on Ontario Highway 407.
- RFID tags are used for electronic toll collection at toll booths with Georgia's Cruise Card, California's FasTrak, Colorado's E-470, Illinois' I-Pass, Oklahoma's Pikepass, the expanding eastern states' E-ZPass system (including Massachusetts's Fast Lane, Delaware, New Hampshire Turnpike, Maryland, New Jersey Turnpike, Pennsylvania Turnpike, West Virginia Turnpike, New York's Thruway system, Virginia, the Maine Turnpike, and Rhode Island's Newport Bridge); Central Florida also utilizes this technology, via its E-PASS System. E-PASS and Sunpass are mutually compatible. Florida's SunPass, various systems in Texas including D/FW's NTTA TollTag, the Austin metro TxTag and Houston HCTRA EZ Tag (which as of early 2007 are all valid on any Texas toll road), Kansas's K-Tag, The "Cross-Israel Highway" (Highway 6), Philippines South Luzon Expressway E-Pass, Brisbane's Queensland Motorways GoVia tag (previously called E-Toll) System in Australia, Autopista del Sol (Sun's Highway), Autopista Central (Central Highway), Autopista Los Libertadores, Costanera Norte, Vespucio Norte Express and Vespucio Sur urban Highways and every forthcoming urban highway (in a "Free Flow" modality) concessioned to private investors in Chile, all toll tunnels in Hong Kong (Autotoll) and all highways in Portugal (Via Verde, the first system in the world to span the entire network of tolls), France (Liber-T system), Italy (Telepass), Spain (VIA-T), Brazil (Sem Parar - Via Fácil). The tags, which are usually the active type, are read remotely as vehicles pass through the booths, and tag information is used to debit the toll amount from a prepaid account. The system helps to speed traffic through toll plazas as it records the date, time, and billing data for the RFID vehicle tag. The plaza- and queue-free 407 Express Toll Route, in the Greater Toronto Area,

allows the use of a transponder (an active tag) for all billing. This eliminates the need to identify a vehicle by license plate.

Public transit (bus, rail, subway)

- Throughout Europe, and in particular in Paris (system started in 1995 by the RATP), Lyon, Bordeaux, Grenoble, Nancy and Marseilles in France, in the whole of the Portuguese highway system and in many Portuguese public car parks, Milan, Turin, Naples and Florence in Italy, and Brussels in Belgium, RFID passes conforming to the Calypso international standard are used for public transport systems. They are also used now in Canada (Montreal), Mexico, Israel, Bogotá and Pereira in Colombia, Stavanger in Norway, Luxembourg, Gävle in Sweden, etc.
- In South Korea, T-money cards can be used to pay for public transit. It can also be used in most convenience stores and vending machines in subways as cash. 90% of cabs in Seoul accept card payment, including most major credit cards and the T-money card. T-money replaced Upass, first introduced for transport payments in 1996 using MIFARE technology.
- In Hong Kong, mass transit is paid for almost exclusively through the use of an RFID technology, called the Octopus Card. Originally it was launched in September 1997 exclusively for transit fare collection, but has grown to be similar to a cash card, and can still be used in vending machines, fast-food restaurants and supermarkets. The card can be recharged with cash at add-value machines or in shops, and can be read several centimetres from the reader. The same applies for Delhi Metro, the rapid transit system in New Delhi, capital city of India.
- In Shanghai the Shanghai Public Transportation Card allows the user to credit money in advance and to be debited according to the distance travelled, as determined by the check-in and check-out stations. The card can also be used to pay taxi drivers, and some shops offer card readers as well.
- The Moscow Metro, the world's second busiest, was the first system in Europe to introduce RFID smartcards in 1998.
- The Washington Metro rail became the first U.S. urban mass-transit system to use RFID technology when it introduced the SmarTrip card in 1999.
- JR East in Japan introduced SUICa (Super Urban Intelligent Card) for transport payment service in its railway transportation service in November 2001, using Sony's FeliCa (Felicity Card) technology. The same Sony technology was used in Hong Kong's Octopus card, and Singapore's EZ-Link card.
- Since 2002, in Taipei, Taiwan the transportation system uses RFID operated cards as fare collection. The EasyCard is charged at local convenience stores and metro

stations, and can be used in Metro, buses and parking lots. The uses are planned to extend all throughout Taiwan in the future.

- In the United States, the Chicago Transit Authority has offered the Chicago Card and the Chicago Card Plus for rail payments across the entire system since 2002 and for bus payments since 2005. The MBTA introduced the RFID enabled CharlieCard across Boston's subway, streetcar, and bus system in 2006, replacing the decades-old token based fare collection system.
- The New York City Metropolitan Transportation Authority conducted an RFID trial that utilized PayPass by MasterCard. The trial primarily took place on the IRT Lexington Avenue Line with several busier stations on other lines also included. The trial ended on May 31, 2009, however the option of using PayPass may be reintroduced on a wider scale at a later date. The MTA is also studying the possibility of accepting SmartLink (introduced by PATH) for fare payment on the New York City Subway and Buses, and as an eventual replacement for the MetroCard.
- In the UK, operating systems for prepaying for unlimited public transport have been devised, making use of RFID technology. The design is embedded in a credit card-like pass, that when scanned reveals details of whether the pass is valid, and for how long the pass will remain valid. The first company to implement this is the NCT company of Nottingham, where the general public affectionately refer to them as "beep cards". It has since been successfully implemented in London, where "Oyster cards" allow for pay-as-you-go travel as well as passes valid for various lengths of time and in various areas.
- In Canada, Metrobus in St. John's adopted RFID on December 1, 2006. In the Greater Toronto and Hamilton areas, under Metrolinx, full implementation of the RFID farecard Presto began in November 2009, and will be rolled out in stages across the network. OC Transpo in Ottawa is also in the process of implementing Presto, with completion expected in late 2011.
- In Oslo, Norway, the upcoming public transport payment is to be entirely RFID-based. The system was slated for introduction around spring 2007.
- The Transperth public transport network in Perth, Western Australia uses RFID technology its SmartRider ticketing system, allowing passengers to "tag on" and "tag off" and be charged automatically, according to how many zones they have travelled.
- In Atlanta, MARTA (Metropolitan Atlanta Rapid Transit Authority) has transitioned its bus and rail lines from coin tokens to the new Breeze Card system which uses RFID tags embedded in disposable paper tickets. More permanent plastic cards are available for frequent users.

- In Rio de Janeiro, "RioCard" passes can be used in buses, ferries, trains and subway. There are two types, one you cannot recharge, the other one can be recharged if it's been bought by the company you work for, if they provided it (only in Brazil).
- In Santiago (Chile) the subway system Metro and the recently implemented public transportation system Transantiago use an RFID card called "Bip" or "Multivia".
- In Medellín (Colombia) the recently-implemented card system for the Metro system uses an RFID card called Cívica.
- In Dubai, (United Arab Emirates) drivers through Sheikh Zayed Road and Garhoud Bridge pay tolls using RFID tags called Salik (road toll). Dubai has also initiated a public transportation card named Nol (which means *fare* in Arabic) for use in the metro, bus, and waterbus. It was introduced to service on 9 September 2009, the day of the official launch of the Dubai Metro.
- In San Diego, California, Metropolitan Transit Systems (MTS), North County Transit District (NCTD), and The San Diego Association Of Governments (SANDAG) use a re-writable RFID smart card referred to locally as the Compass Card, to store daily, weekly, or monthly passes or cash value, making the boarding of buses and trains quicker and simpler.
- In Finland, the RFID travel card system used in the Greater Helsinki region is the largest of systems in Europe that cover all modes of traffic (busses, trams, commuter train units, metros and ferry terminals) operation since 2001. RFID travel card system in Tampere has been in operation since 1995.
- In Cali (Colombia) the recently-implemented card system for the Masivo Integrado de Occidente(MIO) system uses an RFID card.
- In Dublin (Ireland) the LUAS light rail system has been using an RFID enabled 'smart card' system since March 2005.
- In Seattle the Orca Card was introduced in 2009 for fares on buses, ferries, light rail, a street car, and commuter trains. In Tacoma, Washington, a sticker tag is used for paying the toll of the Tacoma Narrows Bridge.
- In Osijek (Croatia) RFID cards have been used to pay for public transportation (buses and trams) since 2008.
- In Buenos Aires (Argentina), Monedero is an RFID card used in all metro lines and, since May 2009, on some bus lines as an experimental program. The card can also be used to pay, as a debit card in some small shops and in toll roads. The Monedero card could be prepaid or linked to a credit card.

- Since 2010, bus transit in Ljubljana (Slovenia) is payable only by RFID with pre-paid city card named Urbana which can be re-filled with monthly passes or cash value on Urbanomats scattered all over the city.
- In Sweden, public transportation has used RFID cards since 2006 in Gothenburg, and since 2009 in Stockholm.

Asset management and retail sales

RFID combined with mobile computing and Web technologies provide a way for organizations to identify and manage their assets. It was initially introduced to major retail by Craig Patterson, Knoxville, TN. Mobile computers, with integrated RFID readers, can now deliver a complete set of tools that eliminate paperwork, give proof of identification and attendance. This approach eliminates manual data entry.

Web based management tools allow organizations to monitor their assets and make management decisions from anywhere in the world. Web based applications now mean that third parties, such as manufacturers and contractors can be granted access to update asset data, including for example, inspection history and transfer documentation online ensuring that the end user always has accurate, real-time data. Organizations are already using RFID tags combined with a mobile asset management solution to record and monitor the location of their assets, their current status, and whether they have been maintained.

RFID is being adopted for item-level retail uses. Aside from efficiency and product availability gains, the system offers a superior form of electronic article surveillance (EAS), and a superior self checkout process for consumers. The first commercial, public item-level RFID retail system installation is believed to be in May 2005 by Freedom Shopping, Inc. in North Carolina, USA.

2009 witnessed the beginning of wide-scale asset tracking with passive RFID. Wells Fargo and Bank of America made announcements that they would track every item in their data centers using passive RFID. Most of the leading banks have since followed suit. The Financial Services Technology Consortium (FSTC) set a technical standard for tagging IT assets and other industries have used that standard as a guideline. For instance the US State Department is now tagging IT assets with passive RFID using the ISO/IEC 18000-6 standard.

Product tracking

RFID use in product tracking applications begins with plant-based production processes, and then extends into post-sales configuration management policies for large buyers.

Casino chip tracking

In 2005, the Wynn Casino, Las Vegas, began placing individual RFID tags on high value chips. These tags allowed casinos the ability to detect counterfeit chips, track betting habits of individual players, speed up chip tallies, and determine counting mistakes of dealers. In 2010, the Bellagio casino was robbed of \$1.5 million in chips. The RFID tags of these chips were immediately invalidated, thus making the cash value of these chips \$0.

IT asset tracking

By 2011 there are more than 100 passive RFID tags that are meant to be specifically mounted on metal. ODIN technologies of Ashburn, Virginia, produced a benchmark which showed varying performance of metal mount tags, with the greatest read distance being just over 25 feet in real-world conditions. In 2010 there are more than 60 metal mount specific RFID tags. Members of the financial service industry, including Wells Fargo, Bank of America, Morgan Stanley, Citigroup, Fidelity and others, are purported to have tagged more than one million assets.

At the same time new integrated circuits (ICs) were introduced by Alien, Impinj and NXP (formerly Philips) which proved much better performance and use of the IT asset tracking application increased. The largest adopter to date appear to be Bank of America and Wells Fargo – each with more than 100,000 assets across more than a dozen data centers.. Some RFID specialists have engineered RFID Softwarespecifically for optimizing RFID performance and scalability.

- High-frequency RFID or HFID/HighFID tags are used in library book or bookstore tracking, jewelry tracking, pallet tracking, building access control, airline baggage tracking, and apparel and pharmaceutical items tracking. High-frequency tags are widely used in identification badges, replacing earlier magnetic stripe cards. These badges need only be held within a certain distance of the reader to authenticate the holder. The American Express Blue credit card now includes a HighFID tag. In Feb 2008, Emirates Airline started a trial of RFID baggage tracing at London and Dubai airports.
- BGN has launched two fully automated Smartstores that combine item-level RFID tagging and SOA to deliver an integrated supply chain, from warehouse to consumer.
- UHF, Ultra-HighFID or UHFID tags are commonly used commercially in case, pallet, and shipping container tracking, and truck and trailer tracking in shipping yards.
- In May 2007, Bear River Supply began utilizing Intellex Corporation's ultrahigh-frequency identification (UHFID) tags to help monitor their agricultural equipment.

- In Colombia, "Federación Nacional de Cafeteros" uses an RFID solution to trace the coffee.
- Purdue Pharma currently uses RFID to track shipments of the painkiller OxyContin.
- In Berlin, Germany, the Berliner Wasserbetriebe (water treatment facility) Uses RFID systems from Psion Teklogix and Elektroniksystem-und-Logistik-GmbH (ESG) to identify and track its 60,000 assets.

Transportation and logistics

- Logistics and transportation are major areas of implementation for RFID technology. For example, yard management, shipping and freight and distribution centers are some areas where RFID tracking technology is used. Transportation companies around the world value RFID technology due to its impact on the business value and efficiency.
- The North American railroad industry operates an automatic equipment identification system based on RFID. Locomotives and rolling stock are equipped with two passive RFID tags (one mounted on each side of the equipment); the data encoded on each tag identifies the equipment owner, car number, type of equipment, number of axles, etc. The equipment owner and car number can be used to derive further data about the physical characteristics of the equipment from the Association of American Railroads' car inventory database and the railroad's own database indicating the lading, origin, destination, etc. of the commodities being carried.
- Aerospace applications that incorporate RFID technology are being incorporated into Network Centric Product Support architecture. This technology serves to help facilitate more efficient logistics support for systems maintenance on-board commercial aircraft.
- Qantas has introduced RFID cards and bag tags in their "Next Generation Check-in" system into the airlines' Australian domestic airport terminals from late 2010 to speed passenger check-in and tracking and recovery of their luggage.
- Baggage passing through the Hong Kong International Airport are individually tagged with "HKIA" RFID tags as they navigate the airport's baggage handling system, which improves efficiency and reduces misplaced items.
- In the Netherlands, the Dutch Government sponsors an RFID project in which Cargobox Europe B.V. tests an intelligent air cargo container. The container has a semi-active, or battery-assisted, tag and can be followed by a series of fixed and handheld readers that will be used in the warehouses of partners throughout the supply chain. This track-and-trace capability is achieved by a web-based IT platform which allows authorized users to view relevant Cargoboxes. The intelligence is in the RFID tags but also in the GPS, GSM, GPRS and airplane detection module that switches off the tag and sensors when the container is airborne.

Animal identification



A sheep with an ear tag.



Animal management using RFID technology. Santa Gertrudis cattle: The calf has an electronic ear tag and *herd management tag* (yellow).

RFID tags for animals represent one of the oldest uses of RFID technology. Originally meant for large ranches and rough terrain, since the outbreak of mad-cow disease, RFID has become crucial in animal identification management.

An implantable variety of RFID tags or transponders can also be used for animal identification. The transponders are more well-known as passive RFID technology, or simply "chips" on animals.

RFID tracking and tracing for meatpackers

The Canadian Cattle Identification Agency began using RFID tags as a replacement for barcode tags. The tags are required to identify a bovine's herd of origin and this is used for tracing when a packing plant condemns a carcass. Currently CCIA tags are used in Wisconsin and by US farmers on a voluntary basis. The USDA is currently developing its own program.

Inventory systems

An advanced automatic identification technology such as the Auto-ID Labs system based on the Radio Frequency Identification (RFID) technology has significant value for inventory systems. Notably, the technology provides an accurate knowledge of the current inventory. In an academic study performed at Wal-Mart, RFID reduced Out-of-

Stocks by 30 percent for products selling between 0.1 and 15 units a day. Other benefits of using RFID include the reduction of labor costs, the simplification of business processes, and the reduction of inventory inaccuracies.

In 2004, Boeing integrated the use of RFID technology to help reduce maintenance and inventory costs on the Boeing 787 Dreamliner. With the high costs of aircraft parts, RFID technology allowed Boeing to keep track of inventory despite the unique sizes, shapes and environmental concerns. During the first six months after integration, the company was able to save \$29,000 in labor.

In 2007, Recall Corporation integrated the use of RFID to help organizations track and audit their records, to support compliance with regulations such as the Sarbanes-Oxley Act and HIPAA.

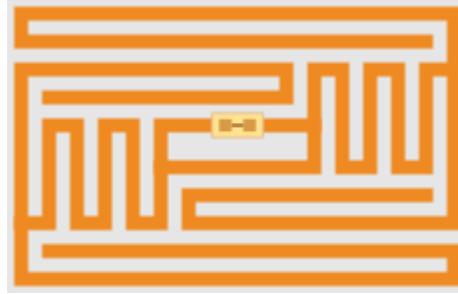
Hospital operating rooms

In 2008, ClearCount Medical introduced the SmartSponge System, the first RFID-based system approved for use in the operating room. The system, consisting of an electronic reader and high frequency RFID-tagged disposable gauze, sponges, and towels, is designed to improve patient safety and O.R. efficiency. The system aims to reduce or eliminate the most common and costly surgical "never event", unintentionally retained foreign objects in surgery. The system automatically provides a device-reconciled count by directly matching the unique identifier on each tagged item both entering into and then out of the surgical case. The system also provides a reusable wand which may be used to scan the patient as an additional safety measure or to assist in locating misplaced sponges.

RFID mandates

Wal-Mart and the United States Department of Defense have published requirements that their vendors place RFID tags on all shipments to improve supply chain management. Due to the size of these two organizations, their RFID mandates impact thousands of companies worldwide. The deadlines have been extended several times because many vendors face significant difficulties implementing RFID systems. In practice, the successful read rates currently run only 80%, due to radio wave attenuation caused by the products and packaging. In time it is expected that even small companies will be able to place RFID tags on their outbound shipments.

Wal-Mart mandate



An EPC RFID tag used by Wal-Mart.

In January 2005, Wal-Mart required its top 100 suppliers to apply RFID labels to all shipments. To meet this requirement, vendors use RFID printer/encoders to label cases and pallets that require EPC tags for Wal-Mart. These smart labels are produced by embedding RFID inlays inside the label material, and then printing bar code and other visible information on the surface of the label.

In October 2005 the University of Arkansas' Information Technology Research Institute released a report on its preliminary study of the impact of RFID on reducing retail out-of-stocks and concluded that RFID reduced OOS by 21% over non-RFID based stores.

Two years later the Wall Street Journal published an article titled "Wal-Mart's Radio-Tracked Inventory Hits Static." The articles stated that the RFID plan set forth by Wal-Mart was "showing signs of fizzling" due to a lack of progress by Wal-Mart executives to introduce the technology to its stores and to the non-existent incentives for suppliers.

In October 2007 Wal-Mart announced new focus areas for its RFID implementation:

- Shipments going to Sam's Club
- Promotional displays and products going to Wal-Mart stores
- Tests to see RFID's impact in improving category management in select areas

Another Wal-Mart division, Sam's Club, has also moved in this direction. It sent letters dated Jan. 7, 2008 to its suppliers, stating that by Jan. 31, 2008, every full single-item pallet shipped to its distribution center in DeSoto, Texas, or directly to one of its stores served by that DC, must bear an EPC Gen 2 RFID tag. Suppliers failing to comply will be charged a service fee.

However, in January 2009 Sam's Club drastically lowered the penalty for failure to tag pallets from \$2 a pallet to just 12 cents a pallet. The 12 cents a pallet is what Wal-Mart estimated it would cost Sam's to do the tagging itself. Sam's also announced that pallet-level tagging is expected to be introduced throughout the entire chain in 2010 while the deadline for tagging individual items was "under review."

In February 2009 Procter & Gamble stated it was ending its promotional program with Wal-Mart after Procter & Gamble "validated" benefits of the RFID program in merchandising and promotional displays. This implied Wal-Mart was not acting on the information to improve store execution.

Department of Defense mandate

The DoD requirements for RFID tags on packages is prescribed in the Defense Federal Acquisition Regulations Supplements (DFARS) 252.211-7006. Positioning of the tag needs to be completed in accordance with the clause and definitions in MIL STD 129 and as of 1 March 2007, EPC Global tags must comply with EPCglobal Class 1 Generation 2 specification.

Promotion tracking

Manufacturers of products sold through retailers promote their products by offering discounts for a limited period on products sold to retailers with the expectation that the retailers will pass on the savings to their customers. However, retailers typically engage in *forward buying*, purchasing more product during the discount period than they intend to sell during the promotion period. Some retailers engage in a form of arbitrage, reselling discounted product to other retailers, a practice known as *diverting*. To combat this practice, manufacturers are exploring the use of RFID tags on promoted merchandise so that they can track exactly which product has sold through the supply chain at fully discounted prices.

Libraries



RFID tags used in libraries: square book tag, round CD/DVD tag and rectangular VHS tag.

Among the many uses of RFID technology is its deployment in libraries. This technology has slowly begun to replace the traditional barcodes on library items (books, CDs, DVDs, etc.). The RFID tag can contain identifying information, such as a book's title or material type, without having to be pointed to a separate database (but this is rare in North America). The information is read by an RFID reader, which replaces the standard barcode reader commonly found at a library's circulation desk. The RFID tag found on library materials typically measures 50×50 mm in North America and 50×75 mm in Europe. It may replace or be added to the barcode, offering a different means of

inventory management by the staff and self service by the borrowers. It can also act as a security device, taking the place of the more traditional electromagnetic security strip.

While there is some debate as to when and where RFID in libraries first began, it was first proposed in the late 1990s as a technology that would enhance workflow in the library setting. Singapore was certainly one of the first to introduce RFID in libraries and Rockefeller University in New York may have been the first academic library in the United States to utilize this technology, whereas Farmington Community Library in Michigan may have been the first public institution, both of which began using RFID in 1999. In Europe, the first public library to use RFID was the one in Hoogezaand-Sappemeer, the Netherlands, in 2001, where borrowers were given an option. To their surprise, 70% used the RFID option and quickly adapted, including elderly people.

Worldwide, in absolute numbers, RFID is used most in the United States (with its 300 million inhabitants), followed by the United Kingdom and Japan. It is estimated that over 30 million library items worldwide now contain RFID tags, including some in the Vatican Library in Rome. At the time of 2010, the largest RFID implementation in academic library is the University of Hong Kong Libraries which have over 1.20 million library items contain RFID tags; whereas the largest implementation for public institution has been installed in Seattle Public Library in the United States.

RFID has many library applications that can be highly beneficial, particularly for circulation staff. Since RFID tags can be read through an item, there is no need to open a book cover or DVD case to scan an item. This could reduce repetitive-motion injuries. Where the books have a barcode on the outside, there is still the advantage that borrowers can scan an entire pile of books in one go, instead of one at a time. Since RFID tags can also be read while an item is in motion, using RFID readers to check-in returned items while on a conveyor belt reduces staff time. But, as with barcode, this can all be done by the borrowers themselves, meaning they might never again need the assistance of staff. Next to these readers with a fixed location there are also portable ones (for librarians, but in the future possibly also for borrowers, possibly even their own general-purpose readers). With these, inventories could be done on a whole shelf of materials within seconds, without a book ever having to be taken off the shelf. In Umeå, Sweden, RFID is being used to assist visually impaired people in borrowing audiobooks. In Malaysia, Smart Shelves are used to pinpoint the exact location of books in Multimedia University Library, Cyberjaya. In the Netherlands, handheld readers are being introduced for this purpose.

The Dutch Union of Public Libraries ('Vereniging van Openbare Bibliotheken') is working on the concept of an interactive 'context library', where borrowers get a reader/headphones-set, which leads them to the desired section of the library (using triangulation methods, rather like GPS) and which they can use to read information from books on the shelves with the desired level of detail (e.g. a section read out loud), coming from the book's tag itself or a database elsewhere, and get tips on alternatives, based on the borrowers' preferences, thus creating a more personalised version of the library. This may also lead them to sections of the library they might not otherwise visit. Borrowers

could also use the system to exchange experiences (such as grading books). This is already done by children in the virtual realm at mijnstempel.nl, but the same could be done in physical form. Borrowers can grade the book at the return desk.

However, as of 2008 this technology remains too costly for many smaller libraries, and the conversion period has been estimated at 11 months for an average-size library. A 2004 Dutch estimate was that a library which lends 100,000 books per year should plan on a cost of €50,000 (borrow- and return-stations: 12,500 each, detection porches 10,000 each; tags 0.36 each). RFID taking a large burden off staff could also mean that fewer staff will be needed, resulting in some of them getting fired, but that has so far not happened in North America where recent surveys have not returned a single library that cut staff because of adding RFID. In fact, library budgets are being reduced for personnel and increased for infrastructure, making it necessary for libraries to add automation to compensate for the reduced staff size. Also, the tasks that RFID takes over are largely not the primary tasks of librarians. A finding in the Netherlands is that borrowers are pleased with the fact that staff are now more available for answering questions.

A concern surrounding RFID in libraries that has received considerable publicity is the issue of privacy. Because RFID tags can — depending on the RFID transmitter & reader — be scanned and read from up to 350 feet or 100 m (eg Smart Label RFID's), and because RFID utilizes an assortment of frequencies (both depending on the type of tag, though), there is some concern over whether sensitive information could be collected from an unwilling source. However, library RFID tags do not contain any patron information, and the tags used in the majority of libraries use a frequency only readable from approximately ten feet. Also, libraries have always had to keep records of who has borrowed what, so in that sense there is nothing new. However, many libraries destroy these records once an item has been returned. RFID would complicate or nullify this respect of readers' privacy. Further, another non-library agency could potentially record the RFID tags of every person leaving the library without the library administrator's knowledge or consent. One simple option is to let the book transmit a code that has meaning only in conjunction with the library's database. Another step further is to give the book a new code every time it is returned. And if in the future readers become ubiquitous (and possibly networked), then stolen books could be traced even outside the library. Tag removal could be made difficult if the tags are so small that they fit invisibly inside a (random) page, possibly put there by the publisher.

Passports

The first RFID passports ("E-passport") were issued by Malaysia in 1998. In addition to information also contained on the visual data page of the passport, Malaysian e-passports record the travel history (time, date, and place) of entries and exits from the country.

Other countries that insert RFID in passports include Norway (2005), Japan (March 1, 2006), most EU countries (around 2006) including Spain, Ireland and the UK, Australia, Hong Kong and the United States (2007), Serbia (July 2008), Republic of Korea (August

2008), Taiwan (December 2008), Albania (January 2009), The Philippines (August 2009), Republic of Macedonia (2010).

Standards for RFID passports are determined by the International Civil Aviation Organization (ICAO), and are contained in ICAO Document 9303, Part 1, Volumes 1 and 2 (6th edition, 2006). ICAO refers to the ISO/IEC 14443 RFID chips in e-passports as "contactless integrated circuits". ICAO standards provide for e-passports to be identifiable by a standard e-passport logo on the front cover.

In 2006, RFID tags were included in new US passports. The US produced 10 million passports in 2005, and it has been estimated that 13 million will be produced in 2006. The chips inlays produced by Smartrac will store the same information that is printed within the passport and will also include a digital picture of the owner. The US State Department initially stated the chips could only be read from a distance of 10 cm (4 in), but after widespread criticism and a clear demonstration that special equipment can read the test passports from 10 meters (33 ft) away, the passports were designed to incorporate a thin metal lining to make it more difficult for unauthorized readers to "skim" information when the passport is closed. The department will also implement Basic Access Control (BAC), which functions as a Personal Identification Number (PIN) in the form of characters printed on the passport data page. Before a passport's tag can be read, this PIN must be entered into an RFID reader. The BAC also enables the encryption of any communication between the chip and interrogator.

Security expert Bruce Schneier has suggested that a mugger operating near an airport could target victims who have arrived from wealthy countries, or a terrorist could design an improvised explosive device which functioned when approached by persons from a particular country if passengers did not put their cards in an area close to their body (high liquid and saline content) or in a foil-lined wallet.

Some other European Union countries are also planning to add fingerprints and other biometric data, while some have already done so.

Schools and universities

School authorities in the Japanese city of Osaka are now chipping children's clothing, back packs, and student IDs in a primary school. A school in Doncaster, England is piloting a monitoring system designed to keep tabs on pupils by tracking radio chips in their uniforms. St Charles Sixth Form College in west London, England, started September, 2008, is using an RFID card system to check in and out of the main gate, to both track attendance and prevent unauthorized entrance. Similarly, Whiteliffe Mount School in Cleckheaton, England uses RFID to track pupils and staff in and out of the building via a specially designed cards. In the Philippines, some schools already use RFID in IDs for borrowing books and also gates in those particular schools have RFID ID scanners for buying items at a school shop and canteen, library and also to sign in and sign out for student and teacher's attendance.

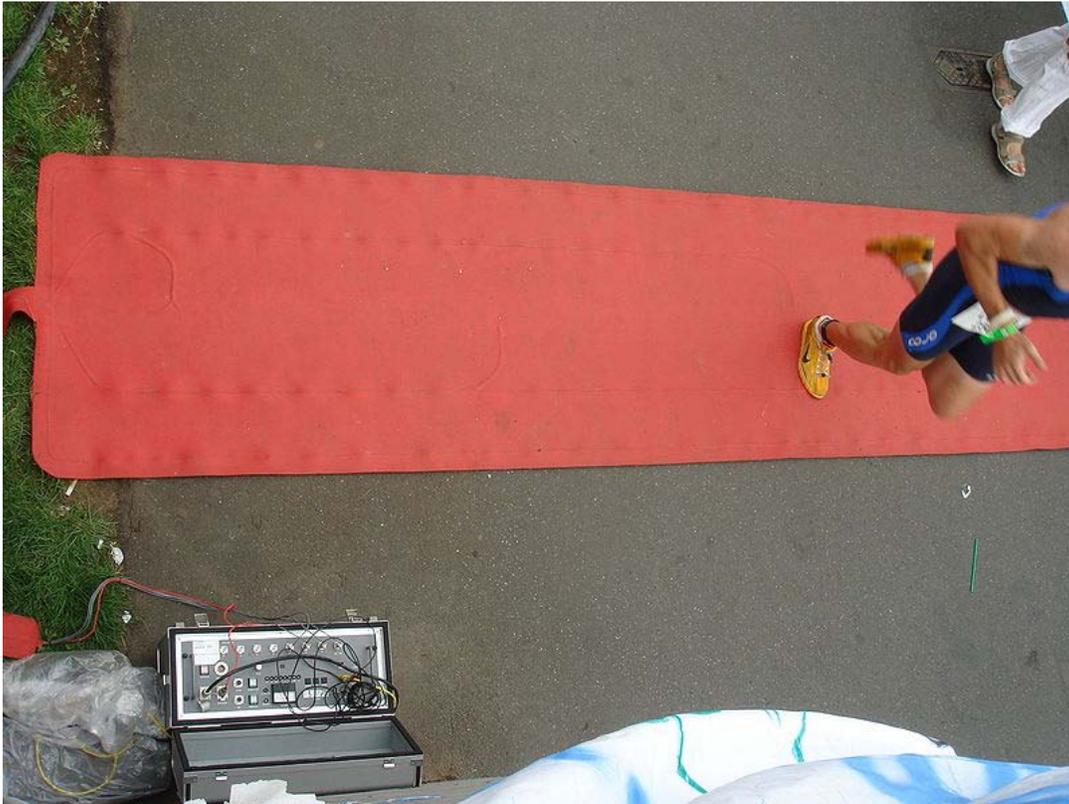
Museums

RFID technologies are now also implemented in end-user applications in museums. An example was the custom-designed temporary research application, "eXspot," at the Exploratorium, a science museum in San Francisco, California. A visitor entering the museum received an RF Tag that could be carried as a card. The eXspot system enabled the visitor to receive information about specific exhibits. Aside from the exhibit information, the visitor could take photographs of themselves at the exhibit. It was also intended to also allow the visitor to take data for later analysis. The collected information could be retrieved at home from a "personalized" website keyed to the RFID tag.

Social retailing

When customers enter a dressing room, the mirror reflects their image and also images of the apparel item being worn by celebrities on an interactive display. A webcam also projects an image of the consumer wearing the item on the website for everyone to see. This creates an interaction between the consumers inside the store and their social network outside the store. The technology in this system is an RFID interrogator antenna in the dressing room and Electronic Product Code RFID tags on the apparel item.

Race timing



J-Chip 8-channel receiver next to timing mat. The athlete wears a chip on a strap around his ankle. Ironman Germany 2007 in Frankfurt.



ChampionChip.

Many forms of RFID race timing have been in use for timing races of different types since the early 1990s. The practice began with pigeon racing, introduced by a company called deister electronic GmbH of Barsinghausen, Germany. It is used for registering race start and end timings for animals or individuals in large running races or multi-sport races where it is impossible to get accurate stopwatch readings for every entrant.

In the race, the racers wear passive or active tags that are read by antennae placed alongside the track or on mats across the track. UHF based tags instead of low or high frequency last-generation tags provide accurate readings with specially designed antennas. Rush error, lap count errors and accidents at start time are avoided since anyone can start and finish any time without being in a batch mode.

Lap scoring

Passive and active RFID systems are used in off-road events such as Orienteering, Enduro and Hare and Hounds racing. Riders have a transponder on their person, normally on their arm. When they complete a lap they swipe or touch the receiver which is connected to a computer and log their lap time. The Casimo Group Ltd sells such a

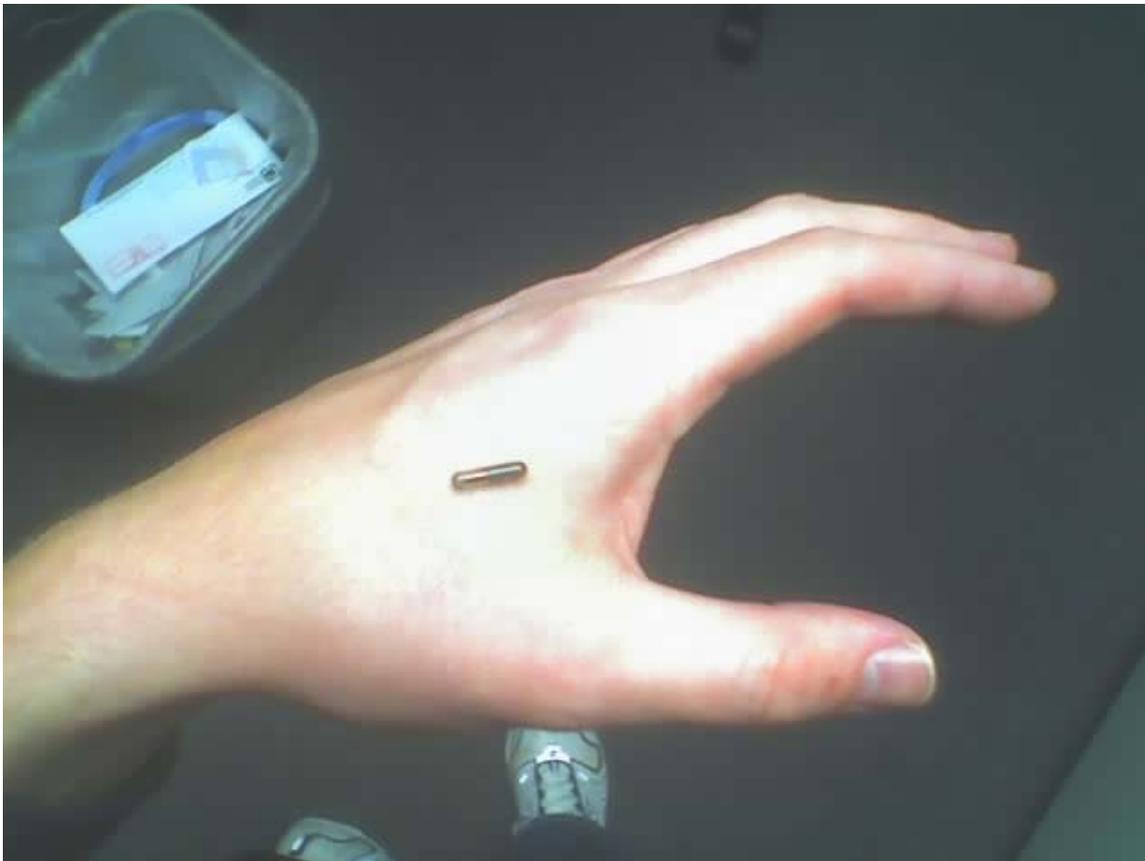
system, as does Sweden's SportIdent and Japan's Micro Talk Systems Corp. which sells the J-Chip system shown in the photo left.

RFID is being adapted by many recruitment agencies which have a PET (Physical Endurance Test) as their qualifying procedure especially in cases where the candidate volumes may run into millions (Indian Railway Recruitment Cells, Police and Power sector).

Ski resorts

A number of ski resorts, particularly in Scandinavia, the French Alps and in the Spanish and French Pyrenees, have adopted RFID tags to provide skiers hands-free access to ski lifts. Skiers do not have to take their passes out of their pockets. The Vail Resorts in Colorado have been using RFID equipped season passes. In 2010, Vail announced that it will be collecting information – vertical feet skied, number of runs taken, lifts used, etc – and all the information will be available to the user online. They are calling this new system EpicMix.

Human implants



Hand with the planned location of the RFID chip.



Just after the operation to insert the RFID tag was completed. The yellow is from the iodine disinfection before inserting the chip.

Implantable RFID chips designed for animal tagging are now being used in humans. An early experiment with RFID implants was conducted by British professor of cybernetics Kevin Warwick, who implanted a chip in his arm in 1998. In 2004 Conrad Chase offered implanted chips in his night clubs in Barcelona and Rotterdam to identify their VIP customers, who in turn use it to pay for drinks.

In 2004, the Mexican Attorney General's office implanted 18 of its staff members with the Verichip to control access to a secure data room.

Security experts have warned against using RFID for authenticating people due to the risk of identity theft. For instance a man-in-the-middle attack would make it possible for an attacker to steal the identity of a person in real-time. Due to the resource constraints of RFIDs it is virtually impossible to protect against such attack models as this would require complex distance-binding protocols.

Privacy advocates have protested against implantable RFID chips, warning of potential abuse and denouncing these types of RFID devices as "spychips", and that use by governments could lead to an increased loss of civil liberties and would lend itself too easily to abuse. One such case of this abuse would be in the microchip's dual use as a

tracking device. Such concerns were justified in the United States, when the FBI program COINTELPRO was revealed to have tracked the activities of high profile political activist and dissident figures. There is also the possibility that the chip's information will be available to those other than governments, such as private business, thus giving employers highly personal information about employees. In addition, privacy advocates state that the information contained in this chip could easily be stolen, so that storing anything private in it would be to risk identity theft.

According to the US Food and Drug Administration (FDA), implantation of an RFID chip poses potential medical downsides. Electrical hazards, MRI incompatibility, adverse tissue reaction, and migration of the implanted transponder are just a few of the potential risks associated with the Verichip ID implant device, according to an October 12, 2004 letter issued by the FDA.

It has been argued that RFID chipping of sex offenders in the US could be politically feasible, and allowable under the U.S. Constitution.

Potential uses

RFID can be used in a variety of applications, such as:

- Access management
- Tracking of goods and RFID in retail
- Tracking of persons and animals
- Toll collection and contactless payment
- Machine readable travel documents
- Smartdust (for massively distributed sensor networks)

- Tracking sports memorabilia to verify authenticity
- Airport baggage tracking logistics

Complement to barcode

RFID tags are often a complement, but not a substitute, for UPC or EAN barcodes. They may not ever completely replace barcodes, due in part to their higher cost and the advantage of multiple data sources on the same object. Also, unlike RFID labels, barcodes can be generated and distributed electronically, *e.g.* via e-mail or mobile phone, for printing and/or display by the recipient. An example is airline boarding passes. The new EPC, along with several other schemes, is widely available at reasonable cost.

The storage of data associated with tracking items will require many terabytes. Filtering and categorizing RFID data is needed to create useful information. It is likely that goods will be tracked by the pallet using RFID tags, and at package level with Universal Product Code (UPC) or EAN from unique barcodes.

The unique identity is a mandatory requirement for RFID tags, despite special choice of the numbering scheme. RFID tag data capacity is large enough that each individual tag will have a unique code, while current bar codes are limited to a single type code for a particular product. The uniqueness of RFID tags means that a product may be tracked as it moves from location to location, finally ending up in the consumer's hands. This may help to combat theft and other forms of product loss. The tracing of products is an important feature that gets well supported with RFID tags containing a unique identity of the tag and also the serial number of the object. This may help companies to cope with quality deficiencies and resulting recall campaigns, but also contributes to concern about tracking and profiling of consumers after the sale.

It has also been proposed to use RFID for POS store checkout to replace the cashier with an automatic system which needs no barcode scanning. In the past this was not possible due to the higher cost of tags and existing POS process technologies. However, Industry Standard, a couture shop and recording studio in Ohio has successfully implemented a POS procedure that allows faster transaction throughput.

An FDA-nominated task force concluded, after studying the various technologies currently commercially available, which of those technologies could meet the pedigree requirements. Amongst all technologies studied including bar coding, RFID seemed to be the most promising and the committee felt that the pedigree requirement could be met by easily leveraging something that is readily available.

Telemetry

Active RFID tags also have the potential to function as low-cost remote sensors that broadcast telemetry back to a base station. Applications of tagometry data could include sensing of road conditions by implanted beacons, weather reports, and noise level monitoring.

Passive RFID tags can also report sensor data. For example, the Wireless Identification and Sensing Platform is a passive tag that reports temperature, acceleration and capacitance to commercial Gen2 RFID readers.

It is possible that active or semi-passive RFID tags used with or in place of barcodes could broadcast a signal to an in-store receiver to determine whether the RFID tag (product) is in the store.

Identification of patients and hospital staff

In July 2004, the US Food and Drug Administration issued a ruling that essentially begins a final review process that will determine whether hospitals can use RFID systems to identify patients and/or permit relevant hospital staff to access medical records. Since then, a number of U.S. hospitals have begun implanting patients with RFID tags and using RFID systems, usually for workflow and inventory management. There is some evidence, as well, that nurses and other hospital staff may be subjected to increased

surveillance of their activities or to labor intensification as a result of the implementation of RFID systems in hospitals. The use of RFID to prevent mixups between sperm and ova in IVF clinics is also being considered.

In October 2004, the FDA approved USA's first RFID chips that can be implanted in humans. The 134 kHz RFID chips, from VeriChip Corp. can incorporate personal medical information and could save lives and limit injuries from errors in medical treatments, according to the company. The FDA approval was disclosed during a conference call with investors. Shortly after the approval, authors and anti-RFID activists Katherine Albrecht and Liz McIntyre discovered an FDA Warning Letter that spelled out serious health risks associated with the VeriChip. According to the FDA, these include "adverse tissue reaction", "migration of the implanted transponder", "failure of implanted transponder", "electrical hazards" and "magnetic resonance imaging [MRI] incompatibility."

St. Clair Hospital in Pittsburgh has deployed an RFID and barcode based bedside medication verification system that improves patient safety by reducing medication errors. Nurses use a PDA equipped with a portable RFID reader and barcode scanner to check patient ID and medications before administering any drugs, including drugs delivered through IV pumps.

To combat home health fraud, the Centers for Medicare & Medicaid Services recently announced heightened scrutiny of the home health care industry. In March, 2009, Elite Medical Supply, a durable medical equipment supplier in New York were one of the first to sign on to combat Medical fraud. They selected CYBRA's EdgeMagic RFID and Bar Code Software to rollout the process.

Regulation and standardization

There is no global public body that governs the frequencies used for RFID. In principle, every country can set its own rules for this. The main bodies governing frequency allocation for RFID are:

- USA: FCC (Federal Communications Commission)
- Canada: Industry Canada - Spectrum Management Branch
- Europe: ERO, CEPT, ETSI, and national administrations (note that the national administrations must ratify the usage of a specific frequency before it can be used in that country)
- Malaysia: Malaysian Communications and Multimedia Commission (MCMC)
- Japan: MIC (Ministry of Internal Affairs and Communications)
- China: Ministry of Information Industry
- Taiwan: NCC (National Communications Commission)
- South Africa: ICASA
- South Korea: Ministry of Knowledge Economy
- Australia: Australian Communications and Media Authority.
- New Zealand: Ministry of Economic Development

- Singapore: Infocomm Development Authority of Singapore
- Brazil: Anatel (Agência Nacional de Telecomunicações)

Low-frequency (LF: 125–134.2 kHz and 140–148.5 kHz) (LowFID) tags and high-frequency (HF: 13.56 MHz) (HighFID) tags can be used globally without a license. Ultra-high-frequency (UHF: 868–928 MHz) (Ultra-HighFID or UHFID) tags cannot be used globally as there is no single global standard. In North America, UHF can be used unlicensed for 902–928 MHz (± 13 MHz from the 915 MHz center frequency), but restrictions exist for transmission power. In Europe, RFID and other low-power radio applications are regulated by ETSI recommendations EN 300 220 and EN 302 208, and ERO recommendation 70 03, allowing RFID operation with somewhat complex band restrictions from 865–868 MHz. Readers are required to monitor a channel before transmitting ("Listen Before Talk"); this requirement has led to some restrictions on performance, the resolution of which is a subject of current research. The North American UHF standard is not accepted in France as it interferes with its military bands. For China and Japan, there is no regulation for the use of UHF. Each application for UHF in these countries needs a site license, which needs to be applied for at the local authorities, and can be revoked. For Australia and New Zealand, 918–926 MHz are unlicensed, but restrictions exist for transmission power.

These frequencies are known as the ISM bands (Industrial Scientific and Medical bands). The return signal of the tag may still cause interference for other radio users.

Some standards that have been made regarding RFID technology include:

- ISO 14223 – Radiofrequency [*sic*] identification of animals – Advanced transponders
- ISO/IEC 14443: This standard is a popular HF (13.56 MHz) standard for HighFIDs which is being used as the basis of RFID-enabled passports under ICAO 9303. The Near Field Communication standard that lets mobile devices act as RFID readers/transponders is also based on ISO/IEC 14443.
- ISO/IEC 15693: This is also a popular HF (13.56 MHz) standard for HighFIDs widely used for non-contact smart payment and credit cards.
- ISO/IEC 18000: Information technology — Radio frequency identification for item management:
 - Part 1: Reference architecture and definition of parameters to be standardized
 - Part 2: Parameters for air interface communications below 135 kHz
 - Part 3: Parameters for air interface communications at 13.56 MHz
 - Part 4: Parameters for air interface communications at 2.45 GHz
 - Part 6: Parameters for air interface communications at 860–960 MHz
 - Part 7: Parameters for active air interface communications at 433 MHz
- ISO/IEC 18092 Information technology — Telecommunications and information exchange between systems — Near Field Communication — Interface and Protocol (NFCIP-1)

- ISO 18185: This is the industry standard for electronic seals or "e-seals" for tracking cargo containers using the 433 MHz and 2.4 GHz frequencies.
- ISO/IEC 21481 Information technology — Telecommunications and information exchange between systems — Near Field Communication Interface and Protocol - 2 (NFCIP-2)
- ASTM D7434, Standard Test Method for Determining the Performance of Passive Radio Frequency Identification (RFID) Transponders on Palletized or Unitized Loads
- ASTM D7435, Standard Test Method for Determining the Performance of Passive Radio Frequency Identification (RFID) Transponders on Loaded Containers
- ASTM D7580 Standard Test Method for Rotary Stretch Wrapper Method for Determining the Readability of Passive RFID Transponders on Homogenous Palletized or Unitized Loads

Groups concerned with standardization are:

- DASH7 Alliance: international industry group formed in 2009 to promote standards and interoperability among extensions to ISO/IEC 18000-7 technologies
- EPCglobal – this is the standardization framework that is most likely to undergo International Standardisation according to ISO rules as with all sound standards in the world, unless residing with limited scope, as customs regulations, air-traffic regulations and others. Currently the big distributors and governmental customers are pushing EPC heavily as a standard well-accepted in their community, but not yet regarded as for salvation to the rest of the world.

EPC Gen2

EPC Gen2 is short for *EPCglobal UHF Class 1 Generation 2*.

EPCglobal (a joint venture between GS1 and GS1 US) is working on international standards for the use of mostly passive RFID and the EPC in the identification of many items in the supply chain for companies worldwide.

One of the missions of EPCglobal was to simplify the Babel of protocols prevalent in the RFID world in the 1990s. Two tag air interfaces (the protocol for exchanging information between a tag and a reader) were defined (but not ratified) by EPCglobal prior to 2003. These protocols, commonly known as Class 0 and Class 1, saw significant commercial implementation in 2002–2005.

In 2004 the Hardware Action Group created a new protocol, the Class 1 Generation 2 interface, which addressed a number of problems that had been experienced with Class 0 and Class 1 tags. The EPC Gen2 standard was approved in December 2004, and is likely to form the backbone of passive RFID tag standards moving forward. This was approved after a contention from Intermec that the standard may infringe a number of their RFID-

related patents. It was decided that the standard itself does not infringe their patents, but that it may be necessary to pay royalties to Intermec if the tag is to be read in a particular manner. The EPC Gen2 standard was adopted with minor modifications as ISO 18000-6C in 2006.

The lowest cost of Gen2 EPC inlay was offered by the now-defunct company SmartCode, at a price of \$0.05 apiece in volumes of 100 million or more. Nevertheless, further conversion (including additional label stock or encapsulation processing/insertion and freight costs to a given facility or DC) and of the inlays into usable RFID labels and the design of current Gen 2 protocol standard will increase the total end-cost, especially with the added security feature extensions for RFID Supply Chain item-level tagging.

Here is the full list of the update on UHF Gen2 Regulation around the world. The list is updated at 2009 January.

Problems and concerns

Data flooding

Each tag generating a message each time when passing a reader may be a desired outcome. However, event filtering is required to reduce this data inflow to a meaningful depiction of moving goods passing a threshold. Various concepts have been designed, mainly offered as *middleware* performing the filtering from noisy and redundant raw data to significant processed data.

Global standardization

The frequencies used for RFID in the USA are currently incompatible with those of Europe or Japan. Furthermore, no emerging standard has yet become as universal as the barcode.

To address international trade concerns, it is necessary to utilize a tag that is operational within all of the international frequency domains. An example of such a tag is a Sentry-M WW from RCD Technology. This mount on metal asset tag provides typical read range of 2 meters (6 ft.). It is functional across the worldwide UHF frequency bands between 860– 960 MHz. It exceeds the Financial Services Technology Consortium RFID Basic Functional Requirements for Data Center Assets in the North American, European and Japanese frequency bands. As a mount on metal solution, the Sentry-M WW is used for tracking many metal assets, such as IT Assets, tools or metal containers.

Security concerns

A primary RFID security concern is the illicit tracking of RFID tags. Tags, which are world-readable, pose a risk to both personal location privacy and corporate/military security. Such concerns have been raised with respect to the United States Department of Defense's recent adoption of RFID tags for supply chain management. More generally,

privacy organizations have expressed concerns in the context of ongoing efforts to embed electronic product code (EPC) RFID tags in consumer products.

EPCglobal Network, by design, is also susceptible to DoS attacks. Using similar mechanism with DNS in resolving EPC data requests, the ONS Root servers become vulnerable to DoS attacks. Any organization planning to embark on EPCglobal Network may cringe upon discovering that the EPCglobal Network infrastructure inherits security weaknesses similar to DNS's.

A second class of defense uses cryptography to prevent tag cloning. Some tags use a form of "rolling code" scheme, wherein the tag identifier information changes after each scan, thus reducing the usefulness of observed responses. More sophisticated devices engage in Challenge-response authentications where the tag interacts with the reader. In these protocols, secret tag information is never sent over the insecure communication channel between tag and reader. Rather, the reader issues a challenge to the tag, which responds with a result computed using a cryptographic circuit keyed with some secret value. Such protocols may be based on symmetric or public key cryptography. Cryptographically-enabled tags typically have dramatically higher cost and power requirements than simpler equivalents, and as a result, deployment of these tags is much more limited. This cost/power limitation has led some manufacturers to implement cryptographic tags using substantially weakened, or proprietary encryption schemes, which do not necessarily resist sophisticated attack. For example, the Exxon-Mobil Speedpass uses a cryptographically-enabled tag manufactured by Texas Instruments, called the Digital Signature Transponder (DST), which incorporates a weak, proprietary encryption scheme to perform a challenge-response protocol for lower cost.

Still other cryptographic protocols attempt to achieve privacy against unauthorized readers, though these protocols are largely in the research stage. One major challenge in securing RFID tags is a shortage of computational resources within the tag. Standard cryptographic techniques require more resources than are available in most low cost RFID devices. RSA Security has patented a prototype device that locally jams RFID signals by interrupting a standard collision avoidance protocol, allowing the user to prevent identification if desired. Various policy measures have also been proposed, such as marking RFID-tagged objects with an industry standard label. RFID security is a very active research field for a few years, with more than 400 scientific papers published since 2002. An extensive list of references in this field can be found at the RFID Security and Privacy Lounge .

Exploitation

Ars Technica reported in March 2006 an RFID buffer overflow bug that could infect airport terminal RFID databases for baggage, and also passport databases to obtain confidential information on the passport holder.

Passports

In an effort to make passports more secure, several countries have implemented RFID in passports. However, the encryption on UK chips was broken in under 48 hours. Since that incident, further efforts have allowed researchers to clone passport data while the passport is being mailed to its owner. Where a criminal used to need to secretly open and then reseal the envelope, now it can be done without detection, adding some degree of insecurity to the passport system.

Shielding

A number of products are available on the market that will allow a concerned carrier of RFID-enabled cards or passports to shield their data. In fact the United States government requires their new employee ID cards to be delivered with an approved shielding sleeve or holder. There are contradicting opinions as to whether aluminum can prevent reading of RFID chips. Some people claim that aluminum shielding, essentially creating a Faraday cage, does work. Others claim that simply wrapping an RFID card in aluminum foil only makes transmission more difficult and is not completely effective at preventing it.

Shielding is again a function of the frequency being used. Low-frequency LowFID tags, like those used in implantable devices for humans and pets, are relatively resistant to shielding, though thick metal foil will prevent most reads. High frequency HighFID tags (13.56 MHz — smart cards and access badges) are sensitive to shielding and are difficult to read when within a few centimetres of a metal surface. UHF Ultra-HighFID tags (pallets and cartons) are difficult to read when placed within a few millimetres of a metal surface, although their read range is actually increased when they are spaced 2–4 cm from a metal surface due to positive reinforcement of the reflected wave and the incident wave at the tag. UHFID tags can be successfully shielded from most reads by being placed within an anti-static plastic bag.

Temperature exposure

Currently, RFID tags are created by gluing an integrated circuit (IC) to an inlay. This poses a problem as vibration and high temperatures will loosen the connection. If the IC loses connection with the inlay, the RFID tag will no longer transmit. A new design was filed for patent (currently pending approval) where the IC is soldered to a circuit board and the circuit board is then soldered to the inlay. This process replaces the adhesive with solder which is much more durable and temperature resistant.

Controversies



Logo of the anti-RFID campaign by German privacy group FoeBuD.

Privacy

The use of RFID technology has engendered considerable controversy and even product boycotts by consumer privacy advocates. Consumer privacy experts Katherine Albrecht and Liz McIntyre are two prominent critics of the technology. The two main privacy concerns regarding RFID are:

- Since the owner of an item will not necessarily be aware of the presence of an RFID tag and the tag can be read at a distance without the knowledge of the individual, it becomes possible to gather sensitive data about an individual without consent.

- If a tagged item is paid for by credit card or in conjunction with use of a loyalty card, then it would be possible to indirectly deduce the identity of the purchaser by reading the globally unique ID of that item (contained in the RFID tag). This is only true if the person doing the watching also had access to the loyalty card data and the credit card data, and the person with the equipment knows where you are going to be.

Most concerns revolve around the fact that RFID tags affixed to products remain functional even after the products have been purchased and taken home and thus can be used for surveillance and other purposes unrelated to their supply chain inventory functions.

The concerns raised by the above may be addressed in part by use of the Clipped Tag. The Clipped Tag is an RFID tag designed to increase consumer privacy. The Clipped Tag has been suggested by IBM researchers Paul Moskowitz and Guenter Karjoth. After the point of sale, a consumer may tear off a portion of the tag. This allows the transformation of a long-range tag into a proximity tag that still may be read, but only at short range – less than a few inches or centimeters. The modification of the tag may be confirmed visually. The tag may still be used later for returns, recalls, or recycling.

However, read range is both a function of the reader and the tag itself. Improvements in technology may increase read ranges for tags. Having readers very close to the tags makes short range tags readable. Generally, the read range of a tag is limited to the distance from the reader over which the tag can draw enough energy from the reader field to power the tag. Tags may be read at longer ranges than they are designed for by increasing reader power. The limit on read distance then becomes the signal-to-noise ratio of the signal reflected from the tag back to the reader. Researchers at two security conferences have demonstrated that passive Ultra-HighFID tags normally read at ranges of up to 30 feet, can be read at ranges of 50 to 69 feet using suitable equipment.



Richard Stallman at WSIS 2005 presenting his RFID badge wrapped with aluminium foil as a way of protesting RFID privacy issues.

In January 2004 privacy advocates from CASPIAN and the German privacy group FoeBuD were invited to the METRO Future Store in Germany, where an RFID pilot project was implemented. It was uncovered by accident that METRO "Payback" customer loyalty cards contained RFID tags with customer IDs, a fact that was disclosed neither to customers receiving the cards, nor to this group of privacy advocates. This happened despite assurances by METRO that no customer identification data was tracked and all RFID usage was clearly disclosed.

During the UN World Summit on the Information Society (WSIS) between the 16th to 18 November 2005, founder of the free software movement, Richard Stallman, protested the use of RFID security cards. During the first meeting, it was agreed that future meetings would no longer use RFID cards, and upon finding out this assurance was broken, he covered his card with aluminum foil, and would only uncover it at the security stations. This protest caused the security personnel considerable concern, with some not allowing him to leave a conference room in which he had been the main speaker, and the prevention of him entering another conference room, where he was due to speak.

In 2004-2005 the Federal Trade Commission Staff conducted a workshop and review of RFID privacy concerns and issued a report recommending best practices.

RFID was one of the main topics of 2006 Chaos Communication Congress (organized by the Chaos Computer Club in Berlin) and triggered a big press debate. Topics included: electronic passports, Mifare cryptography and the tickets for the FIFA World Cup 2006. Talks showed how the first real world mass application of RFID technology at the 2006 FIFA Soccer World Cup worked. Group monochrom staged a special 'Hack RFID' song.

Zeitgeist The Movie presented RFID chips as a negative technology, theorizing that they will one day be used to track the world population and keep them under control.

Human implantation

The Food and Drug Administration in the US has approved the use of RFID chips in humans. Some business establishments give customers the option of using an RFID-based tab to pay for service, such as the *Baja Beach* nightclub in Barcelona. This has provoked concerns into privacy of individuals as they can potentially be tracked wherever they go by an identifier unique to them. There are concerns this could lead to abuse by an authoritarian government or lead to removal of freedoms.

On July 22, 2006, Reuters reported that two hackers, Newitz and Westhues, at a conference in New York City showed that they could clone the RFID signal from a human implanted RFID chip, showing that the chip is not hack-proof as was previously claimed.

Surgery, even on a small scale, comes with its risks. The RFID chip implantation is no exception. According to David B. Smith, the author of "Using Radio Frequency Identification (RFID) Technology in Humans in the United States for Total Control,"

Smith gives the examples of health risks such as "...adverse tissues reaction migration of implanted transponder, compromised information security, failure of implanted transponder, failure of insertion, failure of electronic scanner, electromagnetic interference electrical hazards, magnetic resonance imaging incompatibility, and needle stick" (38). Such risks exist for anyone undergoing an implantation procedure.

Government control

With the rise of technology, some individuals have grown to fear the loss of rights due to RFID human implantation.

By early 2007, Chris Paget of San Francisco, California, showed that RFID information can be pulled from individuals by using only \$250 worth of equipment. This supports the claim that with the information captured, it would be relatively simple to make counterfeit passports.

According to ZDNet, critics believe that this technology will lead to tracking individuals every movement and will be an invasion of privacy. Some conceptualize a future where every movement is tracked by the government. In the book *SpyChips: How Major Corporations and Government Plan to Track Your Every Move with RFID* by Katherine Albrecht and Liz McIntyre, one is encouraged to "imagine a world of no privacy. Where your every purchase is monitored and recorded in a database and your every belonging is numbered. Where someone many states away or perhaps in another country has a record of everything you have ever bought. What's more, they can be tracked and monitored remotely".

Deliberate destruction of RFIDs in clothing and other items

According to an RSA laboratories FAQ, RFID tags can be destroyed by a standard microwave oven; however some types of RFID tags, particularly those constructed to radiate using large metallic antennas (in particular RF tags and EPC tags), may catch fire if subjected to this process for too long (as would any metallic item inside a microwave oven). This simple method cannot safely be used to deactivate RFID features in electronic devices, or those implanted in living tissue, because of the risk of damage to the "host". However the time required is extremely short (a second or two of radiation) and the method works in many other non-electronic and inanimate items, long before thermal buildup (fire) problems become of concern.

Chapter 12

Amateur Radio Repeater



An amateur radio repeater system consisting of a 70cm repeater and a 2 meter digipeater and iGate.

An **amateur radio repeater** is an electronic device that receives a weak or low-level amateur radio signal and retransmits it at a higher level or higher power, so that the signal can cover longer distances without degradation. Many repeaters are located on hilltops or on tall buildings as the higher location increases their coverage area, sometimes referred to as the radio horizon, or "footprint". Amateur radio repeaters are similar in concept to those in use by public safety (police, fire, etc.), business, government, military, and more. Amateur radio repeaters may even use commercially-packaged repeater systems tuned into an amateur radio frequency allocation, but more usually amateur repeaters are assembled from various sources for receivers, transmitters, controllers, power supplies, antennas, and other components.

In amateur radio, repeaters are typically maintained by individual hobbyists or local groups of amateur radio operators. Many repeaters are provided openly to other amateur radio operators and typically not used as a remote base station by a single user or group. In some areas multiple repeaters are linked together to form a wide-coverage network, such as the linked system provided by the Independent Repeater Association which covers most of western Michigan, or the Western Intertie Network System ("WINsystem") that now covers a great deal of California, and is in 17 other states, including Hawaii, along with parts of four other countries, Australia, Canada, Great Britain and Japan.

Services

Services provided by a repeater may include an autopatch connection to a POTS/PSTN telephone line to allow users to make telephone calls from their keypad-equipped radios. These advanced services may be limited to members of the group or club that maintains the repeater. Many amateur radio repeaters typically have a tone access control (CTCSS, CG or PL tone) implemented to prevent them from being keyed-up (operated) accidentally by interference from other radio signals. A few use a digital code system called *DCS*, **DCG** or *DPL* (a Motorola trademark).

In many communities, a repeater has become a major on-the-air gathering spot for the local amateur radio community, especially during "drive time" (the morning or afternoon commuting time). In the evenings local public service nets may be heard on these systems and many repeaters are used by weather spotters. In an emergency or a disaster a repeater can sometimes help to provide needed communications between areas that could not otherwise communicate. Until cellular telephones became popular, it was common for community repeaters to have "drive time" monitoring stations so that mobile amateurs could call in traffic accidents via the repeater to the monitoring station who could relay it to the local police agencies via telephone. Systems with autopatches frequently had (and still have) most of the public safety agencies numbers programmed as speed-dial numbers.

Repeater networks

Repeaters may be linked together in order to form what is known as a *linked repeater system* or *linked repeater network*. In such a system, when one repeater is keyed-up by receiving a signal, all the other repeaters in the network are also activated and will transmit the same signal. The connections between the repeaters are made via radio (usually on a different frequency from the published transmitting frequency) for maximum reliability. Such a system allows coverage over a wide area, enabling communication between amateurs often hundreds of miles (several hundred km) apart. All the user has to know is which channel to use in which area.

In order to get better receive coverage over a wide area, a similar linked setup can also be done with what is known as a *voted receiver system*. In a voted receiver, there are several satellite receivers set up to receive on the same frequency (the one that the users transmit on). All of the satellite receivers are linked to a voting selector panel that switches from receiver to receiver based on the best quieting (strongest) signal, and the output of the selector will actually trigger the central repeater transmitter. A properly adjusted voting system can switch many times a second and can actually "assemble" a multi-syllable word using a different satellite receiver for each syllable. Such a system can be used to widen coverage to low power mobile radios or handheld radios that otherwise would not be able to key up the central location, but can receive the signal from the central location without an issue. Voting systems require no knowledge or effort on the part of the user - the system just seems to have better-than-average handheld coverage.

Repeaters may also be connected to over the Internet using voice over IP (VoIP) techniques. VoIP links are a convenient way to connecting distant repeaters that would otherwise be unreachable by VHF/UHF radio propagation. Popular VoIP amateur radio network protocols include D-STAR, Echolink, IRLP, WIRES and eQSO.

Satellite repeaters

In addition, communications satellites called OSCARs (Orbiting Satellite Carrying Amateur Radio) have been launched with the specific purpose of operating as spaceborne amateur repeaters. The worldwide amateur satellite organization AMSAT designs and builds many of the amateur satellites. Several satellites with amateur radio on board have been designed and built by universities around the world. NASA and AMSAT coordinated the release of SuitSat which was an attempt to make a low cost experimental satellite from a discarded Russian spacesuit outfitted with amateur radio equipment.

Repeater coordination

Having two repeaters operate on the same radio frequency is problematic, as they can interfere with each other, even with selective calling methods enabled. To help minimize this issue, regional repeater coordination organizations have been created. In some jurisdictions, coordination may be required by law or regulation. In others, coordination is done on a voluntary basis, but with a regulatory preference for coordinated repeaters.

USA coordination

In the USA, coordination is optional, but Part 97 rule 205(c) prefers a coordinated repeater over an uncoordinated repeater in disputes over interference. Coordination is overseen by the National Frequency Coordinators' Council (NFCC), a non-profit organization that certifies regional coordinators.

UK repeaters

In the UK, repeaters are managed by the Emerging Technology Co-ordination Committee (ETCC) of the Radio Society of Great Britain and licenced by Ofcom, the industry regulator for communications in the UK.

Equipment



2 Meter GE Mastr II repeater

The most basic repeater consists of an FM receiver on one frequency and an FM transmitter on another frequency usually in the same radio band, connected together so

that when the receiver picks up a signal, the transmitter is keyed and rebroadcasts whatever is heard.

Ham repeaters are found mainly in the VHF six meters (50-54 MHz), two meter (144 - 148 MHz), 220 MHz band (222-224 MHz) and the UHF 70 centimeter (420 - 450 MHz) bands, but can be used on almost any frequency pair above 28 MHz. Note that different countries have different rules; for example, in the United States, the two meter band is 144-148 MHz, while in the United Kingdom (and most of Europe) it's 144-146 MHz.

Repeater frequency sets are known as "repeater pairs," and in the ham radio community most follow *ad hoc* standards for the difference between the two frequencies, commonly called the *offset*. In the USA two-meter band, the standard offset is 600 kHz (0.6 MHz), but some non-conforming *oddball-split* repeaters can be found in various places, and for various reasons, usually technical. The actual frequency pair used is assigned by a local frequency coordinating council.

In the days of crystal-controlled radios, these pairs were identified by the last portion of the transmit (*Input*) frequency followed by the last portion of the receive (*Output*) frequency that the ham would put into the radio. Thus "three-four nine-four" (34/94) meant that hams would transmit on 146.34 MHz and listen on 146.94 MHz (while the repeater would do the opposite, listening on 146.34 and transmitting on 146.94). In areas with many repeaters, "reverse splits" were common (i.e., 94/34), to prevent interference between systems.

Since the late 1970s, the use of synthesized, microprocessor-controlled radios, and widespread adoption of standard frequency splits have changed the way repeater pairs are described. In 1980, a ham might have been told that a repeater was on "22/82" -- today they will most often be told "682 down." The 6 refers to the last digit of 146 MHz, so that the display will read "146.82" (the output frequency), and the radio is set to transmit "down" 600 kHz on 146.22 MHz.

Repeaters typically have a timer to cut off retransmission of a signal that goes too long. Repeaters operated by groups with an emphasis on emergency communications often limit each transmission to 30 seconds, while others may allow three minutes or even longer. The timer restarts after a short pause following each transmission, and many systems feature a beep or chirp tone to signal that the timeout timer has reset.

Simplex repeater

A type of system known as a *simplex repeater* uses a single transceiver and a short-duration voice recorder, which records whatever the receiver picks up for a set length of time (usually 30 seconds or less), then plays back the recording over the transmitter on the same frequency. A common name for them is a "parrot" repeater.

Same-band repeater

Standard repeaters require either the use of two antennas (one each for transmitter and receiver) or a *duplexer* to isolate the transmit and receive signals over a single antenna. The duplexer is a device which prevents the repeater's high power transmitter (on the output frequency) from drowning out the users' signal on the repeater receiver (on the input frequency). A *diplexer* allows two transmitters on different frequencies to use one antenna, and is common in installations where one repeater on 2m and a second on 440 MHz share one feedline up the tower and one antenna.

Most repeaters are remotely controlled through the use of audio tones on a control channel.

Repeaters can be setup as a "Link System" where transmitting on one repeater simultaneously transmits on all repeaters in the system. These systems are used for area or regional communications, for example in Skywarn..

Cross-band repeater

A cross-band repeater (also sometimes called a replexer), is a repeater that retransmits a specific mode on a frequency in one band to a specific mode on a frequency in a different band. This technique allows for smaller size and less complexity of the repeater system. Repeating signals across widely separated bands allows for simple filters to be used to allow one antenna to be used for both transmit and receive at the same time, avoiding the use of complex duplexers to achieve the required rejection for same band repeating. This type of system is used in the OSCAR repeaters.

Contrast with a transponder, which retransmits a range of modes and frequencies from one band to another.

Most dual-band amateur transceivers are capable of cross-band repeat.

Digipeater

Another form of repeater is used in amateur packet radio, a form of digital computer-to-computer communications, and are dubbed "digipeaters" (for *DIGItal rePEATERS*). These repeaters are used for activities and modes such as packet radio, Automatic Position Reporting System, and D-STAR's digital data mode. They operate in a store-and-forward manner, by receiving and then retransmitting on the same frequency after a short delay.

SSTV repeater

An **SSTV repeater** is an amateur radio repeater station for relaying of slow-scan television signals. A typical SSTV repeater is equipped with a HF or VHF transceiver

and a computer with a sound card, which serves as a demodulator/modulator of SSTV signals.

SSTV repeaters are used by amateur radio operators for exchanging pictures. If two stations can not copy each other, they can still communicate through a repeater.

To activate a repeater the station must send a tone of frequency 1750 Hz. Then the repeater is activated and sends K in morse code. The station must start sending a picture in approximately 10 seconds. After reception the received image is transmitted on the repeater's operation frequency.

Repeaters should operate in common SSTV modes, but it depends on the software used (MMSSTV, JVComm32, MSCAN). Some repeater are not activated by audio tone, but instead by the SSTV vertical synchronization signal (VIS code).

When there is no activity on the repeater's frequency, it works as a beacon and periodically send a random picture with identification and a timestamp.

Amateur television repeater

Amateur television (ATV) repeaters are used by amateur radio operators to transmit full motion video. The bands used by ATV repeaters vary by country, but in the US a typical configuration is as a cross-band system with an input on the 33 or 23 cm band and output on 421.25 MHz or, sometimes, 426.25 MHz (70 cm band). These output frequencies happen to be the same as standard cable television channels 57 and 58, meaning that anyone with a cable-ready analog NTSC TV can tune them in without special equipment.

There are also digital amateur TV repeaters that retransmit digital video signals. Frequently DVB-S modulation is used for digital ATV, due to narrow bandwidth needs and high loss tolerances. These DATV repeaters are more prevalent in Europe currently, partially because of the availability of DVB-S equipment.

Transponder

Amateur transponder repeaters are commonly used on amateur satellites. A specified band of frequencies, usually having a bandwidth of 20 to 800 kHz is repeated from one band to another. Transponders may be inverting or non-inverting. An example of an inverting transponder would be a 70cm to 2m transponder which receives on the 432.000 MHz to 432.100 MHz frequencies and transmits on the 146.000 MHz to 146.100 MHz frequencies by inverting the frequency range within the band. In this example, a signal received at 432.001 MHz would be transmitted on 146.099 MHz. Voice signals using USB modulation on the input would result in a LSB modulation on the output, and vice versa.

Operating Terms

Timing Out is a term used to describe the situation where a person talks too long and the repeater timer shuts off the repeater transmitter.

Kerchunking is a term used in ham radio that refers to the act of transmitting a momentary signal to check a repeater without identifying. In many countries, such an act violates amateur radio regulations.

The term "Kerchunk" can also apply to the sound a large Amplitude Modulation Transmitter makes when the operator switches it off and on.

Chapter 13

Magnetic Detector and Radio Receiver Design

Magnetic detector



1902 - 1918: Marconi's Wireless Magnetic Detector (London)

The **magnetic detector** was one of the first practical devices able to make radio signals audible through a pair of headphones. It was not as sensitive as some detectors but despite that, it was favored for early maritime use because of its reliability.

History

During his transatlantic radio communication experiments in December 1902 Marconi found the coherer to be too unreliable/insensitive for detecting the very weak signals inherent in long distance transmissions. It was this need that drove him to develop the "Maggie" or magnetic detector.

The earliest development models and earliest patent of the magnetic detector had a rotating magnet above a stationary segment of iron band with coils on it. It was based on the theory of operation proposed by Rutherford in 1896 (Phil. Trans. Roy. Soc. London A V. 189, pp. 1—24 [1897]). Further developments by Marconi, et al. resulted in a more effective configuration with the moving iron band driven by a clockwork motor and stationary coils.

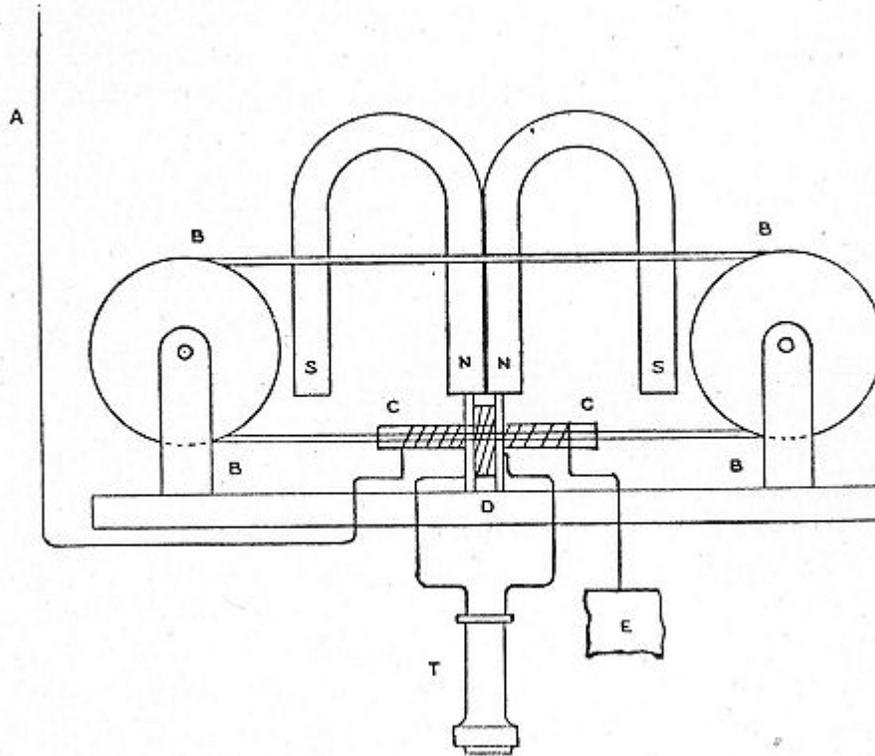
Sir John Ambrose Fleming writes in *The principles of electric wave telegraphy and telephony*:

It was well known long before the middle of the last century that the discharge of a Leyden jar had a magnetizing power. Sir Humphry Davy magnetized sewing-needles with Leyden jar discharges in 1821. Joseph Henry, in the United States, between 1842 and 1850, explored many of the puzzling facts connected with this subject, and only obtained a clue to the anomalies when he realized that the discharge of a condenser through a low resistance circuit is oscillatory in nature. Amongst other things, Henry noticed the power of condenser discharges to induce secondary currents which could magnetize steel needles even when a great distance separated the primary and secondary circuits. He employed this magnetization to test the direction of the secondary currents, and he was followed in the same field of research by Abria, Marianini, Riess, and Matteucci.

In 1870 Lord Rayleigh, in discussing some electromagnetic phenomena, pointed out that the resultant magnetic effect of an oscillatory discharge depends upon the direction of the maximum value of the current during the oscillation, and also that there may be superimposed magnetic effects in the same needle.

In 1895 the subject was again taken up by Professor E. Rutherford, and in a very able paper, published in 1896, he described experiments he had made on the subject.

How it works



A, Antenna Wire; E, Ground-plate; B B ..., Iron band around pulleys; S N ..., Permanent magnets; C C, RF excitation winding on glass tube through which the iron band travels; D, Audio pickup winding; T, Telephone receiver.

The Marconi version consisted of an endless iron band built up of 70 strands of number 40 gage silk-covered iron wire. In operation, the band passes over two grooved pulleys rotated by a wind-up clockwork motor. The iron band passes through the center of glass tube which is close wound with a single layer along several millimeters with number 36 gage silk-covered copper wire. This coil functions as the radio frequency excitation coil. Over this winding is a small bobbin wound with wire of the same gauge to a resistance of about 140 ohms. This coil functions as the audio pickup coil. Around these coils two weak permanent horseshoe magnets are arranged to magnetize the iron band as it passes through the glass tube.

The magnet poles are arranged to create two opposite magnetic fields each directed toward (or away) from the center of the coils in opposite directions along the wire. This functions to magnetize the iron band first in one direction as it approaches the center of the coils, then reverse its magnetism to the opposite direction as it leaves from the other side of the coil. This causes a reversal of the iron band's magnetism just as it passes through the coils. This continuous reversal in magnetism induces a very weak DC current in the audio pickup coil.

The radio signal from the antenna is tuned and passed through the excitation coil, the other end of which is connected to ground. That radio signal in the excitation coil aids the reversal of the magnetization of the iron band as it passes under the audio pickup coil. This works by agitating the magnetic domains to reduce reluctance in the iron band. Changes in the amplitude of the radio signal change the rate of magnetic reversal in the moving iron band. This causes variation in the magnetic flux through the audio pickup coil causing the current in the audio pickup coil to vary. The audio pickup coil is connected to a telephone receiver which converts the varying current to sound.

From a technical standpoint, several subtle prerequisites are necessary for operation. The strength of the magnetic field of the permanent magnets at the iron band must be of the same order of magnitude as the strength of the field generated by the radio frequency excitation coil, allowing the radio frequency signal to significantly contribute to the total the magnetic field. Also, the radio signal applied to the excitation coil roughly matches the impedance of the excitation coil which is very low and requires a special tuner design considerations. The impedance of the telephone earphone must roughly match the impedance of the audio pickup coil, which is a few hundred ohms. The iron band moves a few millimeters per second. Its sensitivity does not approach that of a good mineral detector, yet it is a reliable and surprisingly sensitive detector for radio.

Marconi Company usage

The "Maggie" was the "official" detector used by the Marconi Company in 1902 through 1918, although it was soon supplanted by crystal receivers and multi-element vacuum tubes.

Maintenance

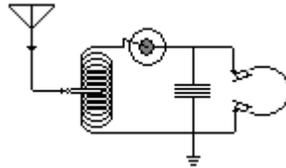
In the *Handbook Of Technical Instruction For Wireless Telegraphists* by: J. C. Hawkhead (Second Edition Revised by H. M. Dowsett) on pp 175 are detailed instructions and specifications for operation and maintenance of the Marconi's Wireless magnetic detector.

Radio receiver design

Radio receiver design includes the electronic design of different components of a radio receiver which processes the radio frequency signal from an antenna in order to produce usable information such as audio. Here we only concentrates on the historical configurations leading up to and including the modern superheterodyne receiver design. The complexity of a modern receiver and the possible range of circuitry and methods employed are more generally covered in electronics and communications engineering. The term *radio receiver* is understood here to mean any device which is intended to receive a radio signal in order to generate useful information from the signal, most notably a recreation of the so-called baseband signal (such as audio) which modulated the radio signal at the time of transmission in a communications or broadcast system.

Crystal radio

A crystal radio uses no active parts: it is powered solely by the radio signal itself, whose detected power feeds headphones in order to be audible at all. In order to achieve even a minimal sensitivity, a crystal radio is limited to low frequencies using a large antenna (usually a long wire). It relies on detection using some sort of semiconductor diode such as the original cat's-whisker diode discovered long before the development of modern semiconductors.



A *crystal set receiver* consisting of an antenna, a variable inductor, a cat's whisker, and a filter capacitor.

- Advantages
 - Simple, easy to make. Here we see a classic design for a clandestine receiver in a POW camp.
- Disadvantages
 - Insensitive, it needs a very strong RF signal and/or a long-wire antenna to operate.
 - Poor selectivity since it only has one tuned circuit.

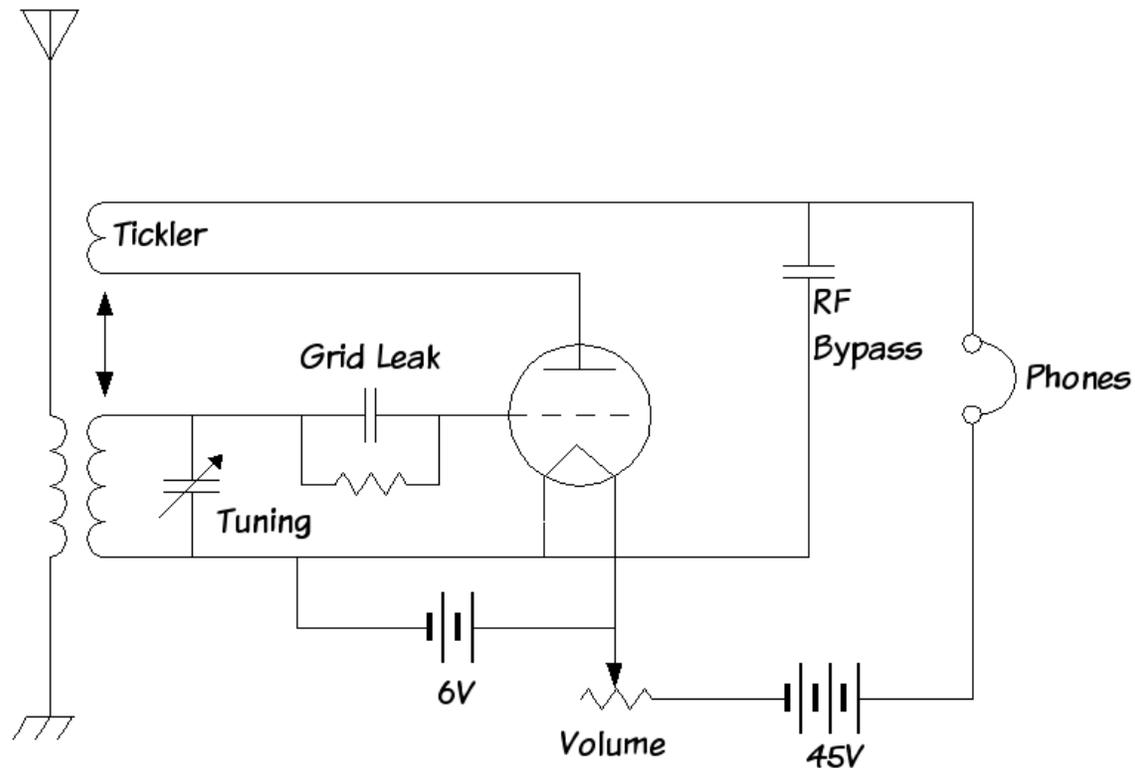
Tuned radio frequency

The tuned radio frequency receiver (TRF) or directly amplifying receiver consists of a radio frequency amplifier having one or more stages all tuned to the desired reception frequency. This is followed by a detector, typically an envelope detector using a diode, followed by audio amplification. This was developed after the invention of the triode vacuum tube, greatly improving the reception of radio signals using electronic amplification which had not previously been available. The greatly improved selectivity of the superheterodyne receiver overtook the TRF design in almost all applications, however the TRF design was still used as late as the 1960's among the cheaper "transistor radios" of that era.

Reflectional

The reflectional receiver was a design from the early 20th century which consists of a single stage TRF receiver but which used the same amplifying tube to also amplify the audio signal after it had been detected. This was in an era where each tube was seen as a major cost (and recipient of electrical power) so that a substantial increase in the number of passive elements would be seen as preferable to including an additional tube. The design tends to be rather unstable, and is obsolete.

Regenerative



Classical regenerative receiver using a single triode vacuum tube. The orientation of the "tickler" coil was carefully adjusted by the operator in order to vary the amount of positive feedback.

The regenerative receiver also had its heyday at the time where adding an active element (vacuum tube) was considered costly. In order to increase the gain of the receiver, positive feedback was used in its single RF amplifier stage; this also increased the selectivity of the receiver well beyond what would be expected from a single tuned circuit. The amount of feedback was quite critical in determining the resulting gain and had to be carefully adjusted by the radio operator. Increasing the feedback beyond a point caused the stage to oscillate at the frequency it was tuned to.

Self-oscillation reduced the quality of its reception of an AM (voice) radio signal but did ironically make it useful as a CW (morse code) receiver inasmuch as the beat signal between the oscillation and the radio signal would produce an audio "beeping" sound. The oscillation of the regenerative receiver could also be an annoying source of local interference. An improved design known as the super-regenerative receiver improved the performance by allowing an oscillation to build up which was then "quenched," with that cycle repeating at a rapid (ultrasonic) rate. From the schematic shown below for a practical regenerative receiver, one can appreciate its simplicity in relation to a multi-stage TRF receiver, while able to achieve the same level of amplification through the use of positive feedback.

Direct conversion

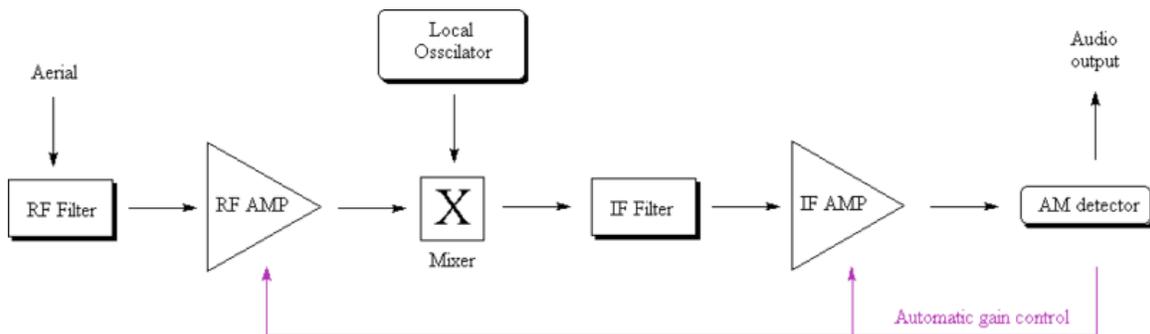
In the Direct conversion receiver, the signals from the antenna are only tuned by a single tuned circuit before entering a mixer where they are mixed with a signal from a local oscillator which is tuned to the carrier wave frequency of the transmitted signal (unlike the superheterodyne design, where the local oscillator is at an offset frequency). The output of this mixer is thus audio frequency, which is passed through a low pass filter into an audio amplifier which may drive a speaker.

For receiving CW (morse code) the local oscillator is tuned to a frequency slightly different from that of the transmitter in order to turn the received signal into an audible "beep."

- Advantages
 - Simpler than a superheterodyne receiver
- Disadvantages
 - Poor rejection of strong signals at adjacent frequencies compared to a superheterodyne receiver.
 - Increased noise or interference when receiving a SSB signal since there is no selectivity against the undesired sideband.

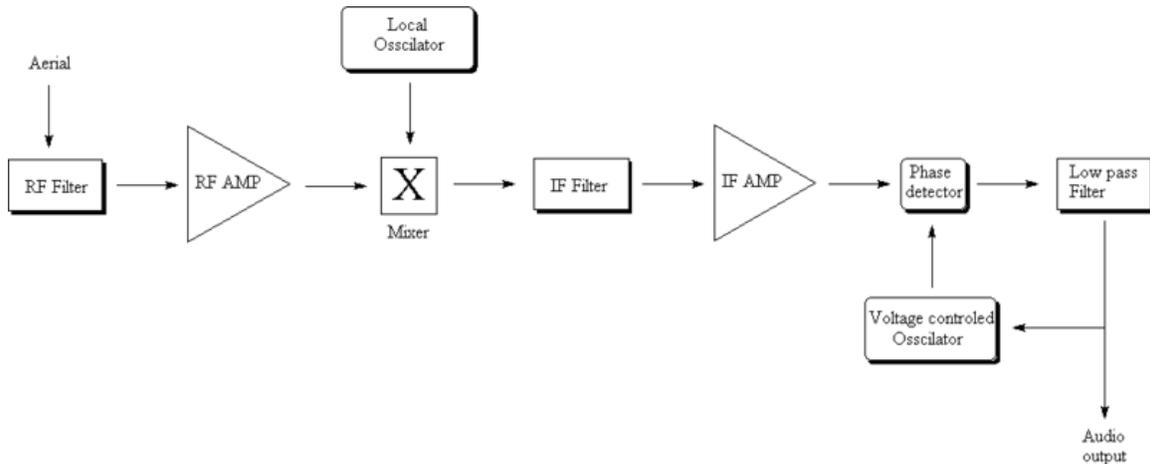
Superheterodyne

Practically all modern receivers are of the superheterodyne design. The RF signal from the antenna may have one stage of amplification to improve the receiver's noise figure, although at lower frequencies this is typically omitted. The RF signal enters a mixer, along with the output of the local oscillator, in order to produce a so-called intermediate frequency (IF) signal. The local oscillator is tuned to a frequency somewhat higher (or lower) than the intended reception frequency so that the IF signal will be at a particular frequency where it is further amplified in a narrow-band multistage amplifier. Tuning the receiver involves changing the frequency of the local oscillator, with further processing of the signal (especially in relation to increasing the selectivity of the receiver) conveniently done at a single frequency (the IF frequency) thus requiring no further tuning for different stations.



A schematic of a superhet AM receiver. Note that the radio includes an AGC loop in order to maintain the RF and IF stages in their linear region, and to produce an audio output not dependent on the signal power received.

Here we show block diagrams for typical superheterodyne receivers for AM and FM broadcast respectively. This particular FM design uses a modern phase locked loop detector, unlike the frequency *discriminator* or ratio detector used in earlier FM receivers.



A schematic of a simple superhet broadcast FM receiver. Note that there is no AGC loop, but simply uses a high-gain IF amplifier which is intentionally driven into saturation (or *limiting*).

For single conversion superheterodyne AM receivers designed for medium wave (AM broadcast) the IF is commonly 455 kHz. Most superheterodyne receivers designed for broadcast FM (88 - 108 MHz) use an IF of 10.7 MHz. TV receivers often use intermediate frequencies of about 40 MHz. Some modern multiband receivers actually convert lower frequency bands first to a much higher frequency (VHF) after which a second mixer with a tunable local oscillator and a second IF stage process the signal as above.