

TRANSMISSION LINES AND WAVEGUIDES

UNIT-I

TRANSMISSION LINE THEORY

INTRODUCTION TO TRANSMISSION LINE THEORY

Transmission Lines and Waveguides

A TRANSMISSION LINE is a device designed to guide electrical energy from one point to another. It is used, for example, to transfer the output rf energy of a transmitter to an antenna. This energy will not travel through normal electrical wire without great losses. Although the antenna can be connected directly to the transmitter, the antenna is usually located some distance away from the transmitter.

On board ship, the transmitter is located inside a radio room, and its associated antenna is mounted on a mast. A transmission line is used to connect the transmitter and the antenna. The transmission line has a single purpose for both the transmitter and the antenna. This purpose is to transfer the energy output of the transmitter to the antenna with the least possible power loss. How well this is done depends on the special physical and electrical characteristics (impedance and resistance) of the transmission line.

TRANSMISSION LINE THEORY

The electrical characteristics of a two-wire transmission line depend primarily on the construction of the line. The two-wire line acts like a long capacitor. The change of its capacitive reactance is noticeable as the frequency applied to it is changed.

Since the long conductors have a magnetic field about them when electrical energy is being passed through them, they also exhibit the properties of inductance. The values of inductance and capacitance presented depend on the various physical factors that we discussed earlier.

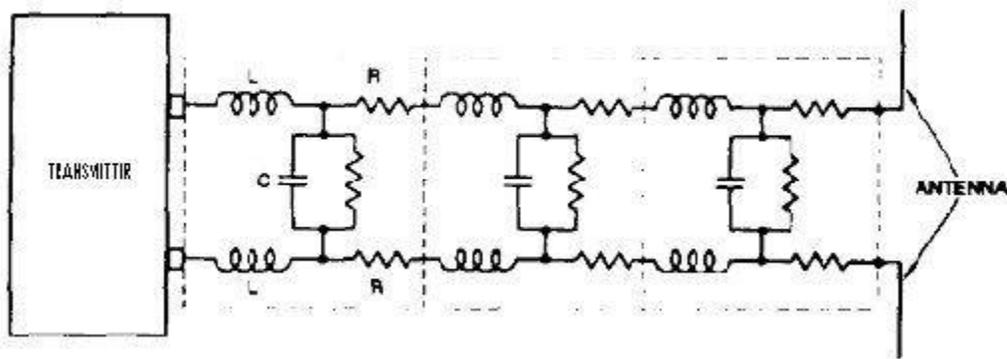
For example, the type of line used, the dielectric in the line, and the length of the line must be considered. The effects of the inductive and capacitive reactance of the line depend on the frequency applied. Since no dielectric is perfect, electrons manage to move from one

conductor to the other through the dielectric.

Each type of two-wire transmission line also has a conductance value. This conductance value represents the value of the current flow that may be expected through the insulation, If the line is uniform (all values equal at each unit length), then one small section of

the line may represent several feet. This illustration of a two-wire transmission line will be used throughout the discussion of transmission lines; but, keep in mind that the principles presented apply to all transmission lines. We will explain the theories using LUMPED CONSTANTS and DISTRIBUTED CONSTANTS to further simplify these principle.

LUMPED CONSTANTS



A transmission line has the properties of inductance, capacitance, and resistance just as the more conventional circuits have. Usually, however, the constants in conventional circuits are lumped into a single device or component. For example, a coil of wire has the property of inductance. When a certain amount of inductance is needed in a circuit, a coil of the proper dimensions is inserted.

The inductance of the circuit is lumped into the one component. Two metal plates separated by a small space, can be used to supply the required capacitance for a circuit. In such a case, most of the capacitance of the circuit is lumped into this one component. Similarly, a fixed resistor can be used to supply a certain value of circuit resistance as a lumped sum. Ideally, a transmission line would also have its constants of inductance, capacitance, and resistance lumped together, as shown in figure 3-1. Unfortunately, this is not the case. Transmission line constants are as described in the following paragraphs.

DISTRIBUTED CONSTANTS

Transmission line constants, called distributed constants, are spread along the entire length of the transmission line and cannot be distinguished separately. The amount of inductance, capacitance, and resistance depends on the length of the line, the size of the conducting wires, the spacing between the wires, and the dielectric (air or insulating medium) between the wires. The following paragraphs will be useful to you as you study distributed

constants on a transmission line.



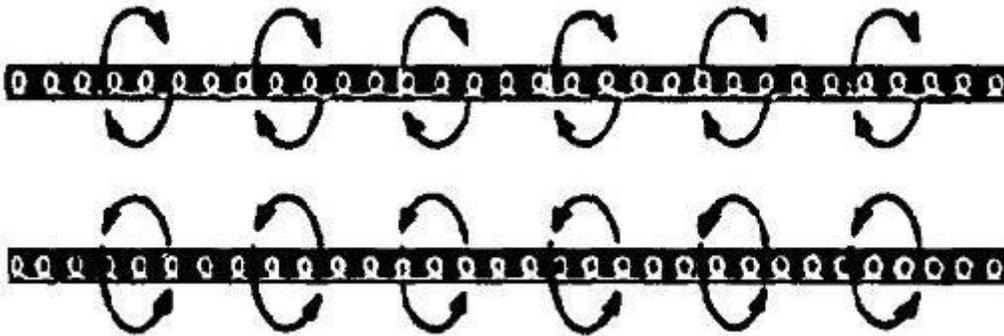
Two-wire transmission line.

Inductance of a Transmission Line

When current flows through a wire, magnetic lines of force are set up around the wire. As the current increases and decreases in amplitude, the field around the wire expands and collapses accordingly. The energy produced by the magnetic lines of force collapsing back into the wire tends to keep the current flowing in the same direction. This represents a certain amount of inductance, which is expressed in microhenrys per unitlength.

Figure illustrates the inductance and magnetic fields of a transmission line.

Capacitance of a Transmission Line



Capacitance also exists between the transmission line wires, as illustrated in figure 3-3. Notice that the two parallel wires act as plates of a capacitor and that the air between them acts as a dielectric. The capacitance between the wires is usually expressed in picofarads per unit length. This electric field between the wires is similar to the field that exists between the two plates of a capacitor.

Resistance of a Transmission Line



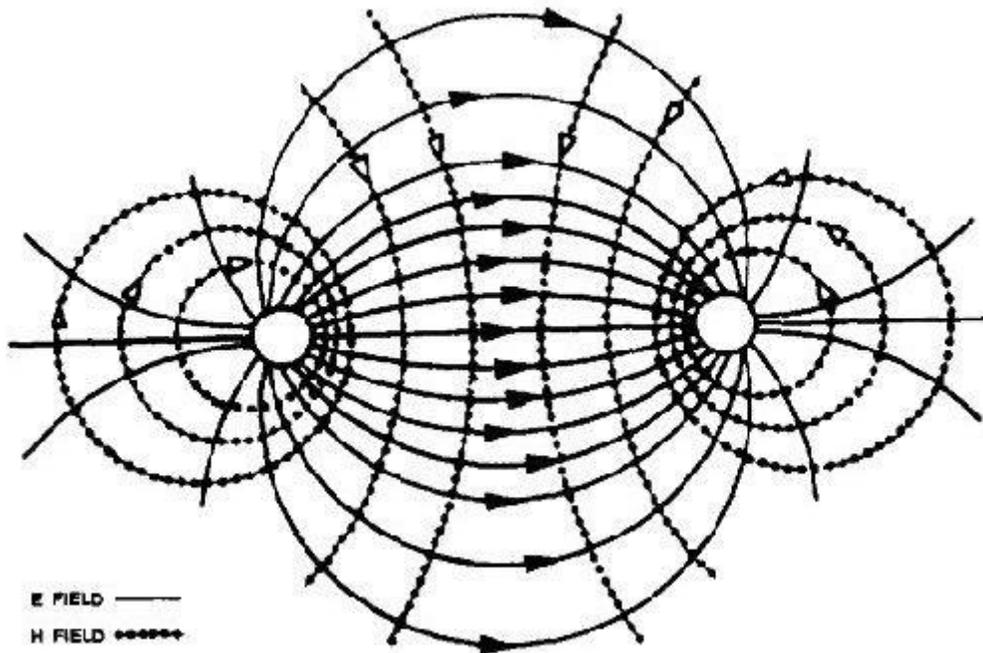
The transmission line shown in figure 3-4 has electrical resistance along its length. This

resistance is usually expressed in ohms per unit length and is shown as existing continuously from one end of the line to the other..

Leakage Current

Since any dielectric, even air, is not a perfect insulator, a small current known as LEAKAGE CURRENT flows between the two wires. In effect, the insulator acts as a resistor, permitting current to pass between the two wires. Figure 3-5 shows this leakage path as resistors in parallel connected between the two lines. This property is called CONDUCTANCE (G) and is the opposite of resistance. Conductance in transmission lines is expressed as the reciprocal of resistance and is usually given in micro mhos per unit length.

ELECTROMAGNETIC FIELDS CHARACTERISTIC IMPEDANCE



The distributed constants of resistance, inductance, and capacitance are basic properties common to all transmission lines and exist whether or not any current flow exists. As soon as current flow and voltage exist in a transmission line, another property becomes quite evident. This is the presence of an electromagnetic field, or lines of force, about the wires of the transmission line.

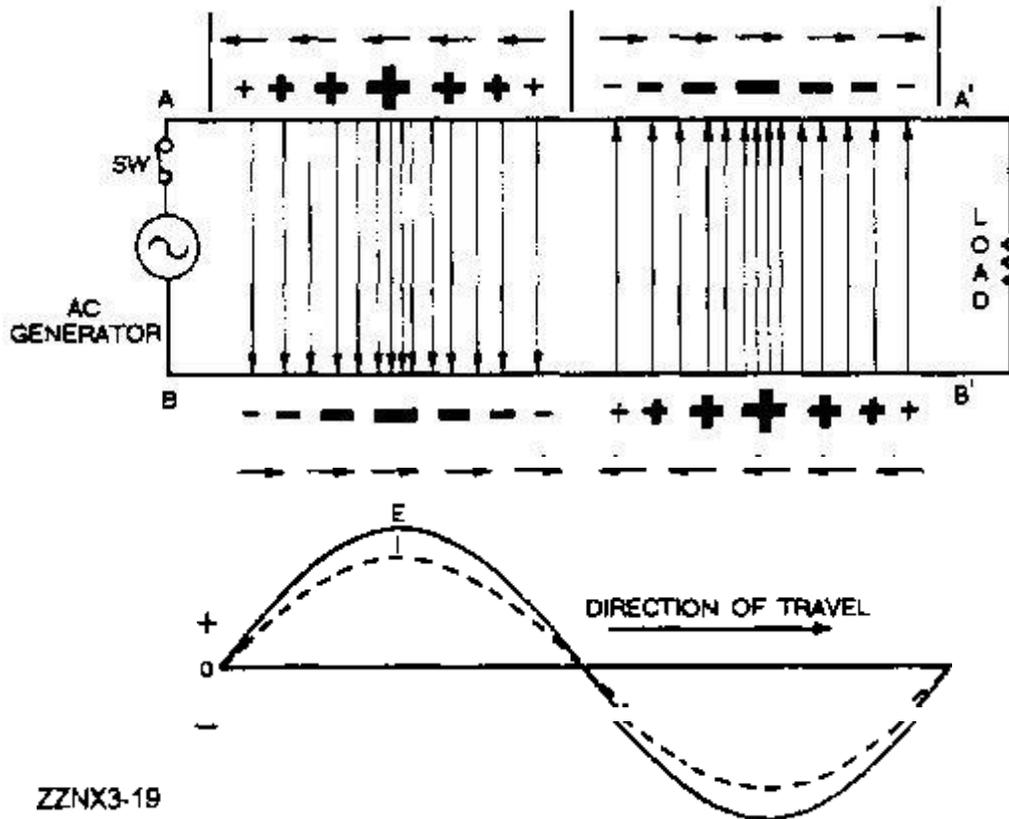
The lines of force themselves are not visible; however, understanding the force that an electron experiences while in the field of these lines is very important to your understanding of energy transmission. There are two kinds of fields; one is associated with voltage and the other with current. The field associated with voltage is called the ELECTRIC (E) FIELD. It exerts a force on any electric charge placed in it. The field associated with current is called a MAGNETIC (H) FIELD, because it tends to exert force on any magnetic pole placed in it. Figure

3-6 illustrates the way in which the E fields and H fields tend to orient themselves between

conductors of a typical two-wire transmission line. The illustration shows a cross section of the transmission lines. The E field is represented by solid lines and the H field by dotted lines. The arrows indicate the direction of the lines of force. Both fields normally exist together and are

spoken of collectively as the electromagnetic Field.

Fields between conductors.



ZZNX3-19

You can describe a transmission line in terms of its impedance. The ratio of voltage to current (E_{in}/I_{in}) at the input end is known as the INPUT IMPEDANCE (Z_{in}). This is the impedance presented to the transmitter by the transmission line and its load, the antenna.

The ratio of voltage to current at the output (E_{OUT}/I_{OUT}) end is known as the OUTPUT IMPEDANCE (Z_{OUT}). This is the impedance presented to the load by the transmission line and its source. If an infinitely long transmission line could be used, the ratio of voltage to current at any point on that transmission line would be some particular value of impedance. This impedance is known as the CHARACTERISTIC IMPEDANCE. The maximum (and most efficient) transfer of electrical energy takes place when the source impedance is matched to the load impedance. This fact is very important in the study of transmission lines and antennas. If the characteristic impedance of the transmission line and the load impedance are equal, energy from the transmitter will travel down the transmission line to the antenna with no power loss caused by reflection.

LINE LOSSES

The discussion of transmission lines so far has not directly addressed LINE LOSSES; actually some losses occur in all lines. Line losses may be any of three types

1. COPPER, DIELECTRIC,
2. RADIATION or INDUCTION LOSSES.

NOTE: Transmission lines are sometimes referred to as RF lines. In this text the terms are used interchangeably.

Copper Losses

One type of copper loss is I^2R LOSS. In RF lines the resistance of the conductors is never equal to zero. Whenever current flows through one of these conductors, some energy is dissipated in the form of heat. This heat loss is a POWER LOSS. With copper braid, which has a resistance higher than solid tubing, this power loss is higher.

Another type of copper loss is due to SKIN EFFECT. When dc flows through a conductor, the movement of electrons through the conductor's cross section is uniform, The situation is somewhat different when ac is applied. The expanding and collapsing fields about each electron encircle other electrons. This phenomenon, called SELF INDUCTION, retards the movement of the encircled electrons.

The flux density at the center is so great that electron movement at this point is reduced. As frequency is increased, the opposition to the flow of current in the center of the wire increases. Current in the center of the wire becomes smaller and most of the electron flow is on the wire surface. When the frequency applied is 100 megahertz or higher, the electron movement in the center is so small that the center of the wire could be removed without any noticeable effect on current. You should be able to see that the effective cross-sectional area decreases as the frequency increases.

Since resistance is inversely proportional to the cross-sectional area, the resistance will increase as the frequency is increased. Also, since power loss increases as resistance increases, power losses increase with an increase in frequency because of skin effect.

Copper losses can be minimized and conductivity increased in an RF line by plating the line with silver. Since silver is a better conductor than copper, most of the current will flow through the silver layer. The tubing then serves primarily as a mechanical support.

Dielectric Losses

DIELECTRIC LOSSES result from the heating effect on the dielectric material between the conductors. Power from the source is used in heating the dielectric. The heat produced is dissipated into the surrounding medium. When there is no potential difference between two conductors, the atoms in the dielectric material between them are normal and the orbits of the electrons are circular.

When there is a potential difference between two conductors, the orbits of the electrons

change. The excessive negative charge on one conductor repels electrons on the dielectric

toward the positive conductor and thus distorts the orbits of the electrons.

A change in the path of electrons requires more energy, introducing a power loss. The atomic structure of rubber is more difficult to distort than the structure of some other dielectric materials.

The atoms of materials, such as polyethylene, distort easily. Therefore, polyethylene is often used as a dielectric because less power is consumed when its electron orbits are

distorted.

RADIATION AND INDUCTION LOSSES

RADIATION and INDUCTION LOSSES are similar in that both are caused by the fields surrounding the conductors. Induction losses occur when the electromagnetic field about a conductor cuts through any nearby metallic object and a current is induced in that object. As a result, power is dissipated in the Object and is lost. Radiation losses occur because some magnetic lines of force about a conductor do not return to the conductor when the cycle alternates. These lines of force are projected into space as radiation, and this result in power losses. That is, power is supplied by the source, but is not available to the load.

VOLTAGE CHANGE

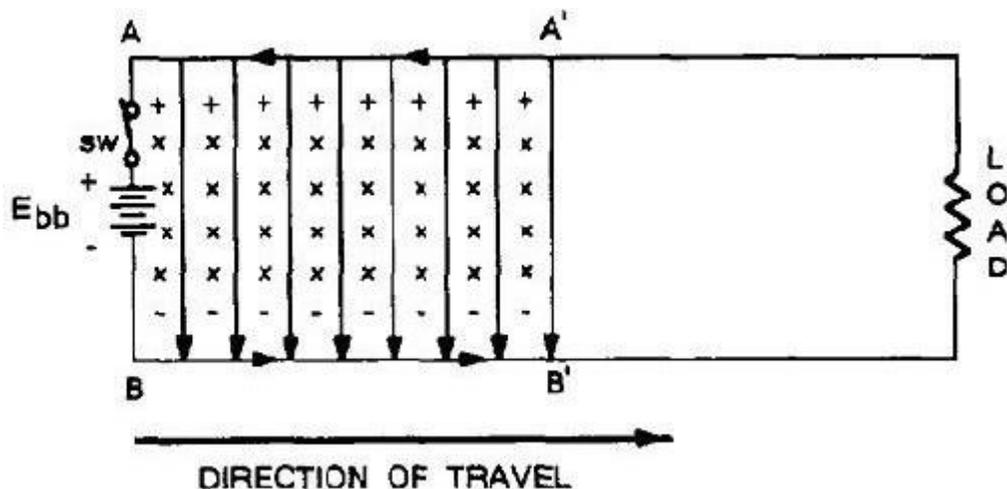
In an electric circuit, energy is stored in electric and magnetic fields. These fields must be brought to the load to transmit that energy. At the load, energy contained in the fields is converted to the desired form of energy

Transmission of Energy

When the load is connected directly to the source of energy, or when the transmission line is short, problems concerning current and voltage can be solved by applying Ohm's law.

When the transmission line becomes long enough so the time difference between a change occurring at the generator and a change appearing at the load becomes appreciable, analysis of the transmission line becomes important.

Dc Applied to a Transmission Line



In figure 3-7, a battery is connected through a relatively long two-wire transmission line to a load at the far end of the line. At the instant the switch is closed, neither current nor voltage exists on the line.

When the switch is closed, point A becomes a positive potential, and point B becomes negative. These points of difference in potential move down the line. However, as the initial points of potential leave points A and B, they are followed by new points of difference in potential, which the battery adds at A and B.

This is merely saying that the battery maintains a constant potential difference between points A and B. A short time after the switch is closed, the initial points of difference in potential have reached points A' and B'; the wire sections from points A to A' and points B to B' are at the same potential as A and B, respectively. The points of charge are represented by plus (+) and minus (-) signs along the wires. The directions of the currents in the wires are represented by the arrowheads on the line, and the direction of travel is indicated by an arrow below the line.

Conventional lines of force represent the electric field that exists between the opposite kinds of charge on the wire sections from A to A' and B to B'. Crosses (tails of arrows) indicate the magnetic field created by the electric field moving down the line. The moving electric field and the accompanying magnetic field constitute an electromagnetic wave that is moving from the generator (battery) toward the load.

This wave travels at approximately the speed of light in free space. The energy reaching the load is equal to that developed at the battery (assuming there are no losses in the transmission line). If the load absorbs all of the energy, the current and voltage will be evenly distributed along the line.

Ac Applied to a Transmission Line

When the battery of figure 3-7 is replaced by an ac generator (fig. 3-8), each successive instantaneous value of the generator voltage is propagated down the line at the speed of light. The action is similar to the wave created by the battery, except the applied voltage is sinusoidal instead of constant. Assume that the switch is closed at the moment the generator voltage is passing through zero and that the next half cycle makes point A positive. At the end of one cycle of generator voltage, the current and voltage distribution will be as shown in figure

In this illustration the conventional lines of force represent the electric fields. For simplicity, the magnetic fields are not shown. Points of charge are indicated by plus (+) and minus (-) signs, the larger signs indicating points of higher amplitude of both voltage and current. Short arrows indicate direction of current (electron flow). The waveform drawn below the transmission line represents the voltage (E) and current (I) waves.

The line is assumed to be infinite in length so there is no reflection. Thus, traveling sinusoidal voltage and current waves continually travel in phase from the generator toward the load, or far end of the line. Waves traveling from the generator to the load are called INCIDENT WAVES. Waves traveling from the load back to the generator are called REFLECTED WAVES and will be explained in later paragraphs.

TRANSMISSION MEDIUMS

The Navy uses many different types of TRANSMISSION MEDIUMS in its electronic

applications. Each medium (line or waveguide) has a certain characteristic impedance value, current-carrying capacity, and physical shape and is designed to meet a particular requirement. The five types of transmission mediums that we will discuss in this topic.

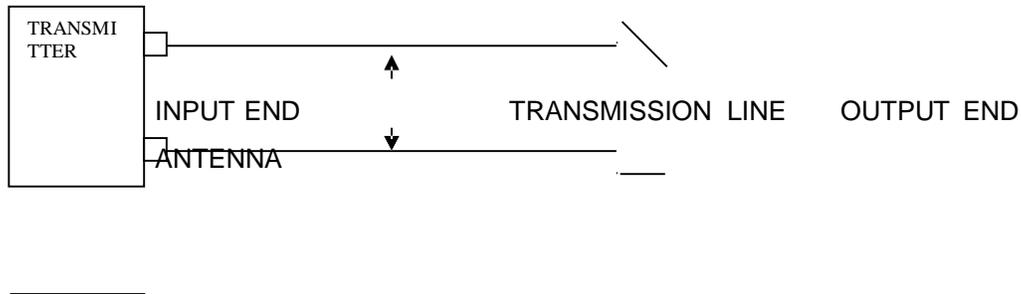
1. PARALLEL-LINE,
2. TWISTED PAIR,

3. SHIELDED PAIR,
4. COAXIAL LINE, and
5. WAVEGUIDES.

The use of a particular line depends, among other things, on the applied frequency, the power-handling capabilities, and the type of installation.

Power Standing-Wave Ratio Parallel Line

The square of the vswr is called the POWER One type of parallel line is the TWO-WIRE OPENSTANDING-WAVE RATIO (pswr). Therefore: LINE, illustrated in figure



This line consists of two wires that are generally spaced from 2 to 6 inches apart by insulating spacers. This type of line is most often used for power lines, rural telephone lines, and telegraph lines. It is sometimes used as a transmission line between a transmitter and an antenna or between a transmitter and a receiver. This ratio is useful because the instruments used to detect standing waves react to the square of the an antenna and a receiver.

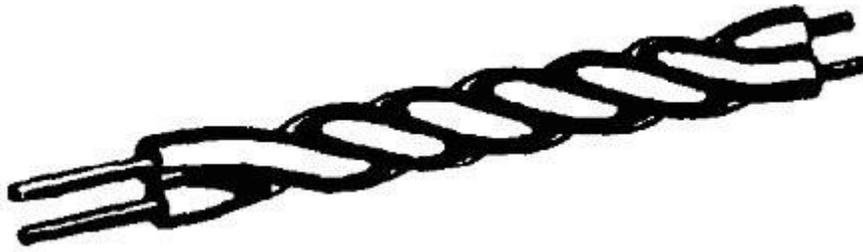
An advantage of this type of line is its simple construction. The principal disadvantages of this type of line are the high radiation losses and electrical noise pickup because of the lack of shielding.

Radiation losses are produced by the changing fields created by the changing current in each conductor. Another type of parallel line is the TWO-WIRE RIBBON (TWIN LEAD) LINE, illustrated in figure 3-10. This type of transmission line is commonly used to connect a television receiving antenna to a home television set.

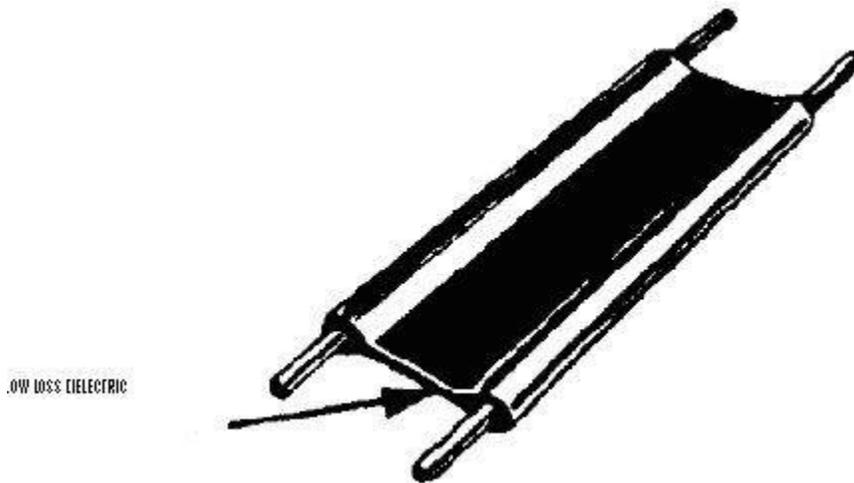
This line is essentially the same as the two-wire open line except that uniform spacing is

assured by embedding the two wires in a low-loss dielectric, usually polyethylene. Since the wires are embedded in the thin ribbon of polyethylene, the dielectric space is partly air and partly polyethylene.

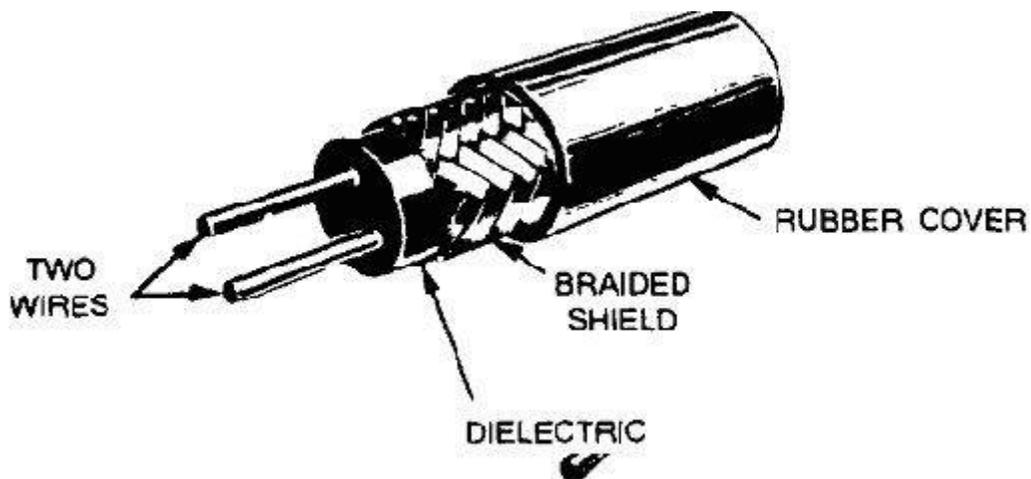
Twisted Pair



The TWISTED PAIR transmission line is illustrated in figure 3-11. As the name implies, the line consists of two insulated wires twisted together to form a flexible line without the use of spacers. It is not used for transmitting high frequency because of the high dielectric losses that occur in the rubber insulation. When the line is wet, the losses increase greatly.



Two-wire ribbon line.



Shielded pair.

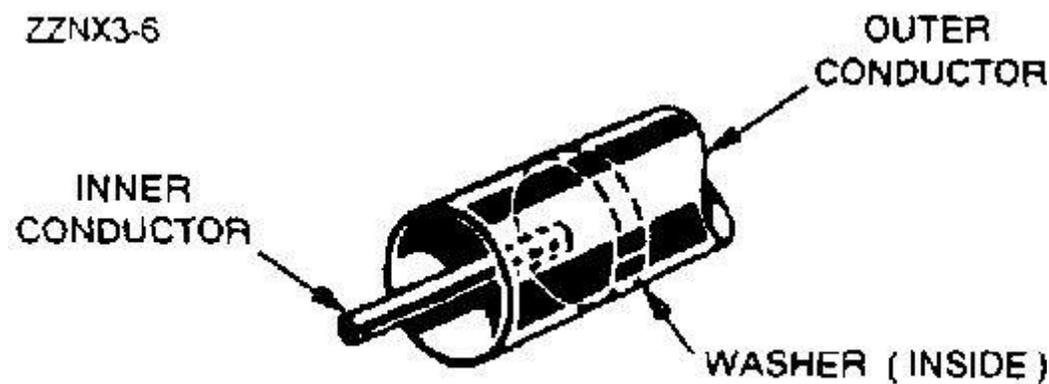
The SHIELDED PAIR, shown in figure, consists of parallel conductors separated from

each other and surrounded by a solid dielectric. The conductors are contained within a braided copper tubing that acts as an electrical shield. The assembly is covered with a rubber or flexible composition coating that protects the line from moisture and mechanical damage. Outwardly, it looks much like the power cord of a washing machine or refrigerator.

Shielded pair.

The principal advantage of the shielded pair is that the conductors are balanced to ground; that is, the capacitance between the wires is uniform throughout the length of the line. This balance is due to the uniform spacing of the grounded shield that surrounds the wires along their entire length. The braided copper shield isolates the conductors from stray magnetic fields.

Coaxial Lines



CABLE WITH WASHER INSULATOR

There are two types of COAXIAL LINES,

1. RIGID (AIR) COAXIAL LINE
2. FLEXIBLE (SOLID)COAXIAL LINE.

The physical construction of both types is basically the same; that is, each contains two concentric conductors.

RIGID CO AXIAL LINE

The rigid coaxial line consists of a central, insulated wire (inner conductor) mounted inside a tubular outer conductor. This line is shown in figure 3-13. In some applications, the inner conductor is also tubular. The inner conductor is insulated from the outer conductor by insulating spacers or beads at regular intervals. The spacers are made of Pyrex, polystyrene, or some other material that has good insulating characteristics and low dielectric losses at high

frequencies.

ADVANTAGES OF RIGID CO AXIAL LINE

The chief advantage of the rigid line is its ability to minimize radiation losses. The electric and magnetic fields in a two-wire parallel line extend into space for relatively great

occur. However, in a coaxial line no electric or magnetic fields extend outside of the outer conductor. The fields are confined to the space between the two conductors, resulting in a perfectly shielded coaxial line. Another advantage is that interference from other lines is reduced.

DIS ADVANTAGES OF RIGID CO AXIAL LINE

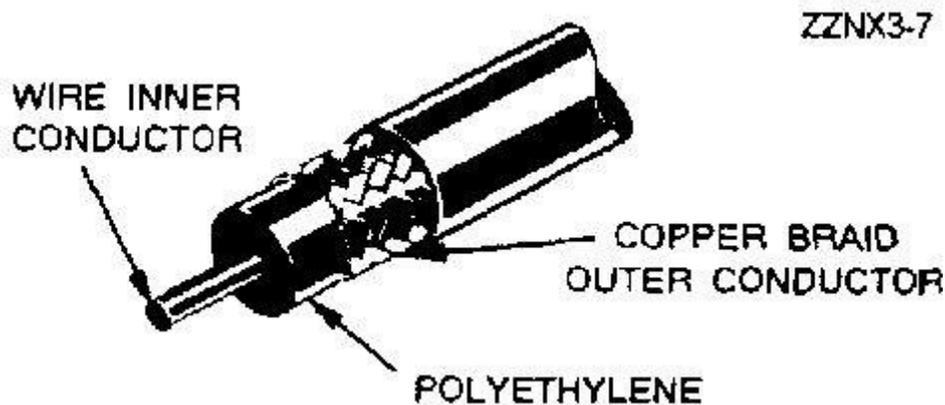
The rigid line has the following disadvantages:

- (1) It is expensive to construct;
- (2) It must be kept dry to prevent excessive leakage between the two Conductors;
- (3) Although high-frequency losses are somewhat less than in previously mentioned

lines, they are still excessive enough to limit the practical length of the line. Leakage caused by the condensation of moisture is prevented in some rigid line applications by the use of an inert gas, such as nitrogen, helium, or argon. It is pumped into the dielectric space of the line at a pressure that can vary from 3 to 35 pounds per square inch. The inert gas is used to dry the line when it is first installed and pressure is maintained to ensure that no moisture enters the line.

FLEXIBLE CO AXIAL LINE

Flexible coaxial lines (fig. 3-14) are made with an inner conductor that consists of flexible wire insulated from the outer conductor by a solid, continuous insulating material. The outer conductor is made of metal braid, which gives the line flexibility. Early attempts at gaining flexibility involved using rubber insulators between the two conductors. However, the rubber insulators caused excessive losses at high frequencies.



Flexible coaxial line.

Because of the high-frequency losses associated with rubber insulators, polyethylene plastic was developed to replace rubber and eliminate these losses. Polyethylene plastic is a solid substance that remains flexible over a wide range of temperatures. It is unaffected by seawater, gasoline, oil, and most other liquids that may be

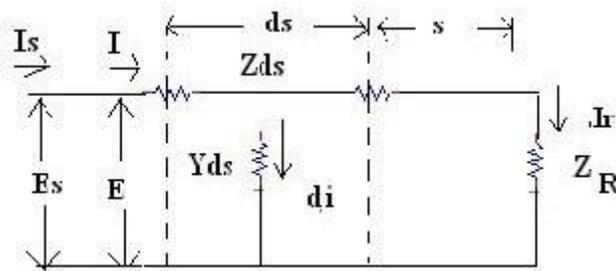
found aboard ship. The use of polyethylene as an insulator results in greater high-frequency

an insulator. However, these losses are still lower than the losses associated with most other solid dielectric Materials.

This concludes our study of transmission lines. The rest of this chapter will be an introduction into the study of waveguides.

THE TRANSMISSION LINE EQUATION – GENERAL SOLUTION

A circuit with distributed parameter requires a method of analysis somewhat different from that employed in circuits of lumped constants. Since a voltage drop occurs across each series increment of a line, the voltage applied to each increment of shunt admittance is a variable and thus the shunted current is a variable along the line.



Hence the line current around the loop is not a constant, as is assumed in lumped constant circuits, but varies from point to point along the line. Differential circuit equations that describes that action will be written for the steady state, from which general circuit equation will be defined as follows.

R = series resistance, ohms per unit length of line(includes both wires)

L = series inductance, henrys per unit length of line

C = capacitance between conductors, faradays per unit length of line

G = shunt leakage conductance between conductors, mhos per unit length

Of line

ωL = series reactance, ohms per unit length of line

$Z = R + j\omega L$

ωC = series susceptance, mhos per unit length of line

$Y = G + j\omega C$

S = distance to the point of observation, measured from the receiving end of the line

I = Current in the line at any point

E = voltage between conductors at any point

l = length of line

The below figure illustrates a line that in the limit may be considered as made up of cascaded infinitesimal T sections, one of which is shown.

This incremental section is of length of ds and carries a current I . The series line impedance being Z ohms and the voltage drop in the length ds is

$$dE = IZds$$

$$\frac{dE}{ds} = IZ \quad (2)$$

The shunt admittance per unit length of line is Y mhos, so that

The admittance of the line is Yds mhos. The current dI that flows across the line or from one conductor to the other is

$$dI = EYds \quad (3)$$

$$\frac{dI}{ds} = EY \quad (4)$$

The equation 2 and 4 may be differentiated with respect to s

$$\frac{d^2 E}{ds^2} = Z \frac{dI}{ds},$$

$$\frac{d^2 I}{ds^2} = Y \frac{dE}{ds}$$

$$\frac{d^2 E}{ds^2} = ZYE \quad (5)$$

$$\frac{d^2 I}{ds^2} = ZYI \quad (6)$$

These are the differential equations of the transmission line, fundamental to circuits of distributed constants.

This results indicates two solutions, one for the plus sign and the other for the minus sign before the radical. The solution of the differential equations are

$$E = Ae^{\sqrt{ZY}s} + Be^{-\sqrt{ZY}s} \quad (7)$$

$$I = Ce^{\sqrt{ZY}s} + De^{-\sqrt{ZY}s} \quad (8)$$

Where A,B,C,D are arbitrary constants of integration.

Since the distance is measured from the receiving end of the line, it is possible to assign conditions such that at

$$s = 0, I = I_R, E = E_R$$

The n equation 7 & 8 becomes

$$\begin{aligned} E_R &= A + B \\ I_R &= C + D \end{aligned} \quad (9)$$

A second set of boundary condition is not available, but the same set may be used over again if a new set of equations are formed by differentiation of equation 7 and 8. Thus

(10)

$$\frac{dE}{ds} = A \sqrt{ZY} e^{\sqrt{ZY}s} - B \sqrt{ZY} e^{-\sqrt{ZY}s} \quad (11)$$

$$IZ = A \sqrt{\frac{Y}{Z}} e^{\sqrt{ZY}s} - B \sqrt{\frac{Y}{Z}} e^{-\sqrt{ZY}s}$$

$$I = A \sqrt{\frac{Y}{Z}} e^{\sqrt{ZY}s} - B \sqrt{\frac{Y}{Z}} e^{-\sqrt{ZY}s}$$

$$\frac{dI}{ds} = C \sqrt{ZY} e^{\sqrt{ZY}s} - D \sqrt{ZY} e^{-\sqrt{ZY}s} \quad (12)$$

$$E = C \sqrt{\frac{Z}{Y}} e^{\sqrt{ZY}s} - D \sqrt{\frac{Z}{Y}} e^{-\sqrt{ZY}s} \quad (13)$$

$$I_R = A \sqrt{\frac{Y}{Z}} - B \sqrt{\frac{Y}{Z}}$$

$$E_R = C \sqrt{\frac{Z}{Y}} - D \sqrt{\frac{Z}{Y}}$$

Simultaneous solution of equation 9 ,12 and 13, along

with the fact that $E_R = I_R Z_R$ and that $\sqrt{\frac{Z}{Y}}$ has

been identified as the Z_0 of the line, leads to solution for the constants of the above equations

as

$$A = \frac{E_R}{2} + \frac{I_R}{2} \sqrt{\frac{Z}{Y}} = \frac{E_R}{2} + \frac{Z_0}{Z_R}$$

$$B = \frac{E_R}{2} - \frac{I_R}{2} \sqrt{\frac{Z}{Y}} = \frac{E_R}{2} - \frac{Z_0}{Z_R}$$

$$C = \frac{I_R}{2} + \frac{E_R}{2} \sqrt{\frac{Y}{Z}} = \frac{I_R}{2} + \frac{Z_R}{Z_0}$$

$$D = \frac{I_R}{2} - \frac{E_R}{2} \sqrt{\frac{Y}{Z}} = \frac{I_R}{2} - \frac{Z_R}{Z_0}$$

The solution of the differential equations of the transmission line may be written

$$E = \frac{E_R}{2} \left(1 + \frac{Z_R}{Z_0} \right) e^{\sqrt{ZY}s} + \frac{E_R}{2} \left(1 - \frac{Z_R}{Z_0} \right) e^{-\sqrt{ZY}s}$$

$$I = \frac{I_R}{2} \left(1 + \frac{Z_R}{Z_0} \right) e^{\sqrt{ZY}s} - \frac{I_R}{2} \left(1 - \frac{Z_R}{Z_0} \right) e^{-\sqrt{ZY}s} \quad (14)$$

The above equations are very useful form for the voltage and current at any point on a transmission line.

After simplifying the above equations we get the final and very useful form of equations for voltage and current at any point on a transmission line, and are solutions to the wave equation

$$E = E_R \cosh \sqrt{ZY} s + I_R Z_0 \sinh \sqrt{ZY} s$$

$$I = I_R \cosh \sqrt{ZY} s + \frac{E_R}{Z} \sinh \sqrt{ZY} s$$

Wavelength, Velocity of propagation

Wavelength

The distance the wave travels along the line while the phase angle is changed through 2π radians is called wavelength.

$$\lambda = 2\pi / \beta$$

The change of 2π in phase angle represents one cycle in time and occurs in a distance of one wavelength,

$$\lambda = v/f$$

Velocity

$$V = f \lambda$$

$$V = \omega / \beta$$

This is the velocity of propagation along the line based on the observation of the change in the phase angle along the line. It is measured in miles/second if β is in radians per meter.

We know that

$$Z = R + j \omega L$$

$$Y = G + j \omega C$$

Then

$$\begin{aligned} \gamma = \alpha + j \beta &= \sqrt{ZY} \\ &= \sqrt{RG - \omega^2 LC + j \omega (LG + CR)} \end{aligned}$$

Squaring on both sides

$$\alpha^2 + 2j\alpha\beta - \beta^2 = RG - \omega^2 LC + j\omega(LG + CR)$$

Equating real parts and imaginary parts we get

$$\alpha = \sqrt{\frac{RG - \omega^2 LC + \sqrt{(RG - \omega^2 LC)^2 + \omega^2 (LG + CR)^2}}{2}}$$

And the equation for β is

$$\beta = \sqrt{\frac{\omega^2 LC - RG + \sqrt{(RG - \omega^2 LC)^2 + \omega^2 (LG + CR)^2}}{2}}$$

In a perfect line $R=0$ and $G=0$, Then the above equation would be

$$\beta = \omega \sqrt{LC}$$

And the velocity of propagation for such an ideal line is given by

$$v = \frac{\omega}{\beta}$$

Thus the above equation showing that the line parameter values fix the velocity of propagation.

DISTORTION

Wave-form distortion

The value of the attenuation constant α has been determined that

$$\alpha = \sqrt{\frac{RG - \omega^2 LC + \sqrt{(RG - \omega^2 LC)^2 + \omega^2 (LG + \omega C)}}{2}}$$

In general α is a function of frequency. All the frequencies transmitted on a line will then not be attenuated equally. A complex applied voltage, such as voice voltage containing many frequencies, will not have all frequencies transmitted with equal attenuation, and the received for will be identical with the input waveform at the sending end. This variation is known as frequency distortion.

Phase Distortion

The of propagation has been stated that

$$\beta = \sqrt{\frac{\omega^2 LC - RG + \sqrt{(RG - \omega^2 LC)^2 + \omega^2 (LG + \omega C)}}{2}}$$

It is apparent that α and β do not both involve frequency in same manner and that the velocity of propagation will in general be some function of frequency.

All the frequencies applied to a transmission line will not have the same time of transmission, some frequencies delayed more than the others. For an applied voice voltage waves the received waves will not be identical with the input wave form at the receiving end, since some components will be delayed more than those of the other frequencies. This phenomenon is known as **delay or phase distortion**.

Frequency distortion is reduced in the transmission of high quality radio broadcast programs over wire line by use of equalizers at line terminals

These circuits are networks whose frequency and phase characteristics are adjusted to be inverse to those of the lines, resulting in an over all uniform frequency response over the desired frequency band.

Delay distortion is relatively minor importance to voice and music transmission because of the characteristics of ear. It can be very serious in circuits intended for picture transmission, and applications of the coaxial cable have been made to overcome the difficulty.

In such cables the internal inductance is low at high frequencies because of skin effect, the resistance small because of the large conductors, and capacitance and leakage are small because of the use of air dielectric with a minimum spacers. The velocity of propagation is raised and made more nearly equal for all frequencies.

The distortion less line

If a line is to have neither frequency nor delay distortion, then attenuation constant and velocity

of propagation cannot be function of frequency.

We know that

$$v = \frac{c}{\beta}$$

Then the phase constant be a direct function of frequency.

$$\beta = \sqrt{\frac{R^2 LC - RG + \sqrt{(RG - R^2 LC)^2 + R^2 (LG + CR)^2}}{2}}$$

The above equation shows that if the the term under the second radical be reduced to equal

$$(RG + R^2 LC)^2$$

Then the required condition for β is obtained. Expanding the term under the internal radical and forcing the equality gives

$$R^2 G^2 - 2R^2 LCRG + R^4 L^2 C^2 + R^2 L^2 G^2 + 2R^2 LCRG + R^2 C^2 = (RG + R^2 LC)^2$$

This reduces to

$$- 2R^2 LCRG + R^2 L^2 G^2 + R^2 C^2 R^2 = 0$$

$$(LG - CR)^2 = 0$$

Therefore the condition that will make phase constant a direct form of frequency is

$$LG = CR$$

A hypothetical line might be built to fulfill this condition. The line would then have a value of β obtained by use of the above equation.

Already we know that the formula for the phase constant

$$\beta = \sqrt{LC}$$

Then the velocity of propagation will be

$$v = 1/\sqrt{LC}$$

This is the same for the all frequencies, thus eliminating the delay distortion.

We know that the equation for attenuation constant

$$\alpha = \sqrt{\frac{RG - R^2 LC + \sqrt{(RG - R^2 LC)^2 + R^2 (LG + CR)^2}}{2}}$$

May be made independent of frequency if the term under the internal radical is forced to reduce to

$$(RG + R^2 LC)^2$$

Analysis shows that the condition for the distortion less line $LG = CR$, will produce the desired result, so that it is possible to make attenuation constant and velocity independent of frequency simultaneously. Applying the condition $LG = CR$ to the expression for the attenuation gives

$$\alpha = \sqrt{RG}$$

This is the independent of frequency, thus eliminating frequency distortion on a line. To achieve

this condition

$$LG = CR$$

$$\frac{L}{C} = \frac{R}{G}$$

Require a very large value of L, since G is small. If G is intentionally increased, α and attenuation are increased, resulting in poor line efficiency.

To reduce R raises the size and cost of the conductors above economic limits, so that the hypothetical results cannot be achieved.

THE TELEPHONE CABLE

In the ordinary telephone cable the wires are insulated with paper and twisted in pairs. This construction results in negligible values of inductance and conductance so that reasonable assumptions in the audio range of frequencies are that

$$Z = R$$

$$Y = j\omega C$$

We know that $\gamma = \sqrt{RG - \omega^2 LC + j\omega(LG + CR)}$

$$\gamma = j\omega CR =$$

$$\alpha = \sqrt{\frac{\omega^2 CR}{2}}$$

With $L=0$, this equation becomes $\beta = \sqrt{\frac{\omega^2 CR}{2}}$

$$v = \frac{\omega}{\beta} = \sqrt{\frac{2}{\omega^2 CR}}$$

It should be observed that both α and the velocity are functions of frequency, such that the higher frequencies are attenuated more and travel faster than the lower frequencies. Very considerable frequency and delay distortion is the result on the telephone cable.

INDUCTANCE LOADING OF TELEPHONE CABLE

A distortion less line with distributed parameters suggest a remedy for the severe frequency and delay distortion experienced on long cables. It was indicated that it was necessary the L/C ratio to achieve distortion less conditions. Heaviside suggested that the inductance be increased,

And Pupin developed the theory that made possible this increase in the inductance by **LUMPED INDUCTORS** spaced at regular intervals along the line. This use of inductance is called

In some submarine cables, distributed or uniform loading is obtained by winding the

cable with a high permeability steel tape such as permalloy. This method is employed because of the practical difficulties of designing lumped loading coils for such underwater circuits.

uniformly loaded cable circuit for which it may be assumed that $G = 0$ and for which L has been increased so that ωL is large with respect to R . Then

$$Z = R + j\omega L$$

$$Y = j\omega C$$

Since,

$$Z = \sqrt{R^2 + \omega^2 L^2} \angle \frac{\pi}{2} - \tan^{-1} \frac{R}{\omega L}$$

Then

$$\gamma = \sqrt{YZ}$$

$$= \sqrt{\sqrt{R^2 + \omega^2 L^2} \angle \frac{\pi}{2} - \tan^{-1} \frac{R}{\omega L} \times j\omega C \angle \frac{\pi}{2}}$$

$$= \omega \sqrt{LC} \sqrt{1 + \frac{R^2}{\omega^2 L^2}} \angle \frac{\pi}{2} - \frac{1}{2} \tan^{-1} \frac{R}{\omega L}$$

In view of the fact that R is small with respect to ωL , the term $\frac{R^2}{\omega^2 L^2}$ may be dropped, and the propagation constant becomes

$$\gamma = \omega \sqrt{LC} \angle \frac{\pi}{2} - \frac{1}{2} \tan^{-1} \frac{R}{\omega L}$$

$$\text{If } \theta = \frac{\pi}{2} - \frac{1}{2} \tan^{-1} \frac{R}{\omega L}$$

$$\cos \theta = \cos\left(\frac{\pi}{2} - \frac{1}{2} \tan^{-1} \frac{R}{\omega L}\right) = \sin \angle \frac{1}{2} \tan^{-1} \frac{R}{\omega L}$$

For a small angle

$$\sin \theta = \tan \theta = \theta$$

$$\cos \theta = \frac{R}{\omega L}$$

Finally the propagation constant may be written as,

$$\gamma = \omega \sqrt{LC} (\cos \theta + j \sin \theta) = \omega \sqrt{LC} \left[\frac{R}{\omega L} + j \right]$$

Therefore, for the uniformly loaded cable,

$$\alpha = \frac{R}{2} \sqrt{\frac{L}{C}}$$
$$\beta = \frac{1}{\sqrt{LC}}$$
$$v = \frac{1}{\beta} = \sqrt{LC}$$

It is readily observed that, under the assumptions of $G=0$ and ωL large with respect to R ,

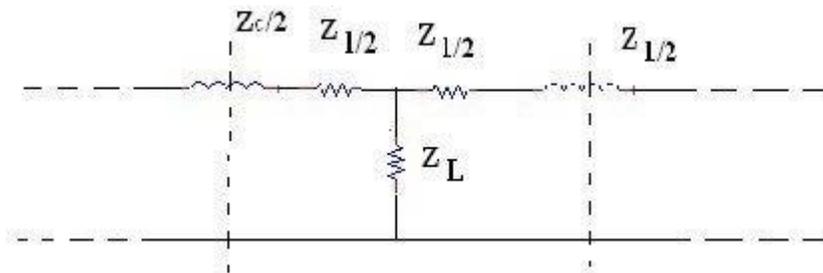
the

attenuation and velocity are both independent of frequency and the loaded cable will be distortion less. The expression for attenuation constant shows that the attenuation may be reduced by increasing L , provided that R is not also increased too greatly. Continuous or uniform loading is expensive and achieves only a small increase in L per unit length.

Lumped loading is ordinarily preferred as a means of transmission improvement for cables. The improvement obtainable on open wire line is usually not sufficient to justify the extra cost of the loading inductors.

CAMPBELL'S EQUATION

An analysis for the performance of a line loaded at uniform intervals can be obtained by considering a symmetrical section of line from the center of one loading coil to the center of the next, where the loading coil of the inductance is Z_c .



The section line may be replaced with an equivalent T section having symmetrical series arms. Adopting the notation of filter circuits one of these series arms is called $Z_1/2$ and is

$$\frac{Z_1}{2} = Z_0 \tanh \frac{N\gamma}{2}$$

Where N is the number miles between loading coils and γ is the propagation constant per mile. Upon including half a loading coil, the equivalent series arm of the loaded section becomes

$$\frac{Z_1'}{2} = \frac{Z_c}{2} + \tanh \frac{N\gamma}{2}$$

The shunt z_2 arm of the equivalent T section is

$$Z_2 = \frac{Z_0}{\sinh N\gamma}$$

An equation relating that γ and the circuit element of a T section was already derived, which

may be applied to the loaded T section as

$$\cosh N\gamma' = 1 + \frac{Z_1'}{2Z_2}$$

$$= 1 + \frac{Z_0}{\sinh N\gamma}$$

By use of exponential it can be shows that

$$\tanh \frac{N\gamma}{2} = \frac{\cosh N\gamma - 1}{\sinh N\gamma}$$

So that the above equation reduces to

$$\cosh N\gamma' = \frac{Z_c}{2Z_0} \sinh N\gamma + \cosh N\gamma$$

This expression is known as **Campbell's** Equation and permits the determination of a value for γ' of a line section consisting partially of lumped and partially of distributed elements.

Campbell's equation makes possible the calculation of the effects of loading coils in reducing attenuation and distortion on lines.

For a cable Z2 of the above figure is essentially capacitive and the cable capacitance plus lumped inductances appear similar to the circuit of the **low pass filter**

It is found that for frequencies below the cutoff, given by

$$f_0 = \frac{1}{\pi\sqrt{LC}}$$

The attenuation is reduced as expected, but above cutoff the attenuation rises as a result of filter action. This cutoff frequency forms a definite upper limit to successful transmission over cables.

It can be raised by reducing L but this expedient allows the attenuation to rise.

The cutoff frequency also be reduced by spacing the closer together, thus reducing C and more closely approximating the distributed constant line, but the cost increases rapidly.

In practice, a truly distortion less line is not obtained by loading, because R and L are to some extent functions of frequency. Eddy current losses in the loading inductors aggravate this condition. However, a major improvement is obtained in the loaded cable for a reasonable frequency range.

INPUT IMPEDANCE AND TRANSFER IMPEDANCE

The input impedance of a transmission line has already been obtained as

$$Z_s = Z_0 = \frac{Z_R \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_R \sinh \gamma l}$$

In terms of exponentials, this is

$$Z_S = Z_0 \frac{e^{\gamma l} + Ke^{\gamma l}}{e^{\gamma l} - Ke^{\gamma l}}$$

If the voltage at the sending –end terminals is known, it is convenient to have the transfer

received current can be computed directly. The sending end voltage E_s is

$$E_s = \frac{E_R (Z_R + Z_0)}{2Z_0} (e^{\gamma l} + K e^{-\gamma l})$$

$$I_s = \frac{I_R (Z_R + Z_0)}{2} (e^{\gamma l} + K e^{-\gamma l})$$

For which the transfer impedance is

$$Z_T = \frac{E_s}{I_s} = \frac{(Z_R + Z_0)}{2} (e^{\gamma l} + K e^{-\gamma l})$$

By substituting for K, The above equation becomes

$$Z_T = Z_R \frac{e^{\gamma l} + e^{-\gamma l}}{2} + \frac{e^{\gamma l} - e^{-\gamma l}}{2}$$

This is recognizable as

$$Z_T = Z_R \cosh \gamma l + Z_0 \sinh \gamma l$$

If the expression is desired in terms of the hyperbolic functions.

Open and short circuited lines

As limited cases it is convenient to consider lines terminated in open circuit or short circuit, that is with $Z_R = \infty$ or $Z_R = 0$. The input impedance of a line of length l is

$$Z_s = Z_0 \frac{Z_R \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_R \sinh \gamma l}$$

And for the short circuit case $Z_R = 0$, so that

$$Z_s = Z_0 \tanh \gamma l$$

Before the open circuit case is considered, the input impedance should be written

$$Z_s = Z_0 \frac{\cosh \gamma l + (Z_0 / Z_R) \sinh \gamma l}{(Z_0 / Z_R) \cosh \gamma l + \sinh \gamma l}$$

The input impedance of the open circuited line of length l, with $Z_R = \infty$, is

$$Z_{oc} = Z_0 \coth \gamma l$$

By multiplying the above two equations it can be seen that

$$Z_0 = \sqrt{Z_{oc} Z_{sc}}$$

This is the same result as was obtained for a lumped network. The above equation supplies a

www.padeepz.net

very valuable means of experimentally determining the value of z_0 of a line.

Also fro the same two equations

$$\tanh \gamma l = \sqrt{\frac{Z_{sc}}{Z_{oc}}}$$

$$\gamma l = \tanh^{-1} \sqrt{\frac{Z_{sc}}{Z_{oc}}}$$

Use of this equation in experimental work requires the determination of the hyperbolic tangent of a complex angle. If

$$\tanh \gamma l = \tanh(\alpha + j\beta)l = U + jV$$

Then it can be shown that

$$\begin{aligned} \tanh 2\alpha l &= \frac{2U}{1+U^2+V^2} \\ \text{and} \quad \tanh 2\beta l &= \frac{2V}{1-U^2-V^2} \end{aligned}$$

The value of β is uncertain as to quadrant. Its proper value may be selected if the approximate velocity of propagation is known.

REFLECTION FACTOR AND REFLECTION LOSS

REFLECTION FACTOR

$$K = \frac{|2\sqrt{Z_1 Z_2}|}{|Z_1 + Z_2|}$$

The term K denotes the reflection factor. This ratio indicates the change in current in the load due to reflection at the mismatched junction and is called the reflection factor.

REFLECTION LOSS

Reflection loss is defined as the number of nepers or decibels by which the current in the load under image matched conditions would exceed the current actually flowing in the load. This reflection loss involves the reciprocal of the reflection factor K.

$$\text{Reflection loss, nepers} = \ln \left| \frac{Z_1 + Z_2}{2\sqrt{Z_1 Z_2}} \right|$$

$$\text{Reflection loss, db} = 20 \log \left| \frac{Z_1 + Z_2}{2\sqrt{Z_1 Z_2}} \right|$$

INSERTION LOSS

Insertion loss of the line or network is defined as the number of nepers or decibels by which the current in the load is changed by the insertion.

T AND π EQUIVALENT TO LINES

The design of an equivalent T section from measurement on a network. These relations were

$$Z_1 = Z_{1oc} - \sqrt{Z_{2oc}(Z_{1oc} - Z_{2oc})}$$

$$Z_2 = Z_{2oc} - \sqrt{Z_{2oc}(Z_{1oc} - Z_{2oc})}$$

$$Z_3 = \sqrt{Z_{2oc}(Z_{1oc} - Z_{2oc})}$$

The input impedance of open circuited and short circuited lines were already developed.

$$Z_{1oc} = \frac{Z_0}{\tanh \gamma l} = Z_0 \frac{e^{\gamma l} + e^{-\gamma l}}{e^{\gamma l} - e^{-\gamma l}}$$

$$Z_{1sc} = Z_0 \tanh \gamma l = Z_0 \frac{e^{\gamma l} - e^{-\gamma l}}{e^{\gamma l} + e^{-\gamma l}}$$

Since a line is symmetrical network,

$$Z_{1oc} = Z_{2oc}$$

The Z_3 or shunt element of a T section that will be equivalent, in so far as external voltages

and currents are connected, to the long line can then be readily obtained as

$$Z_3 = \sqrt{\frac{Z_0}{\tanh \gamma l} - Z_0 \tanh \gamma l}$$

$$= \frac{Z_0}{\sinh \gamma l}$$

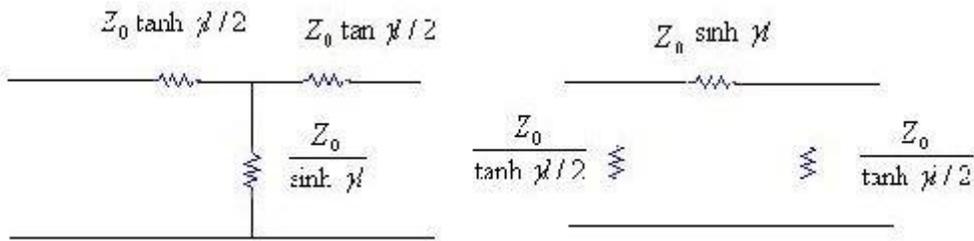
The series elements for the equivalent section then are

$$Z_1 = Z_2 = Z_{1oc} = Z_3 = \frac{Z_0 (e^{\gamma l} + e^{-\gamma l})}{e^{\gamma l} - e^{-\gamma l}} - \frac{Z_0}{e^{\gamma l} - e^{-\gamma l}}$$

$$= Z_0 \frac{(e^{\gamma l/2} + e^{-\gamma l/2})^2}{(e^{\gamma l/2} - e^{-\gamma l/2})(e^{\gamma l/2} + e^{-\gamma l/2})}$$

$$Z_1 = Z_2 = Z_0 \tanh \gamma l / 2$$

The T – section equivalent for the long line, made up of these elements, is shown in the below figure. It is useful in certain types of line calculations.



A π - section equivalent for the line may likewise be determined from the terminal measurements.

Because of symmetry,

$$\begin{aligned}
 Z_A = Z_C &= \frac{Z_{2oc} Z_{1sc}}{Z_{2oc} \sqrt{Z_{2oc} (Z_{1oc} - Z_{1sc})}} \\
 &= \frac{Z_0^2}{\square e^{\gamma l} + e^{-\gamma l} \diamond 2Z_0} \\
 &\quad \square Z_0 \diamond e^{\gamma l} - e^{-\gamma l} \diamond e^{\gamma l} - e^{-\gamma l}
 \end{aligned}$$

Therefore

$$Z_A = Z_C = \frac{Z_0}{\tanh(\gamma l / 2)}$$

The Z_B arm simply obtained as

$$Z_B = \frac{Z_{2oc} Z_{1sc}}{\sqrt{Z_{2oc} (Z_{1oc} - Z_{1sc})}}$$

$$Z_B = \frac{Z_0^2}{Z_0 / \sinh \gamma l} = Z_0 \sinh \gamma l$$

UNIT –II

HIGH FREQUENCY TRANSMISSION LINES

INTRODUCTION

When a line, either open wire or coaxial, is used at frequencies of a megacycle or more, it is found that certain approximations may be employed leading to simplified analysis of line performance. The assumptions usually made are:

1. Very considerable skin effect, so that currents may be assumed as flowing on a conductor surfaces, internal inductance then being zero.
2. That $\omega L \gg R$ when computing Z . This assumption is justifiable because it is found that the resistance increases because of skin effect with \sqrt{f} while the line resistance increases directly with f .

3. The lines are well enough constructed that G may be considered zero

The analysis is made in either of two ways, depending on whether R is merely small with respect to ωL or R is small, the line is considered completely negligible compared with ωL . If R is small, the line is considered one of small dissipation, and this concept is useful when lines are employed as circuit elements or where resonance properties are involved. If losses were neglected then infinite current or voltages would appear in calculations, and physical reality would not be achieved.

In applications where losses may be neglected, as in transmission of power at high efficiency, R may be considered as negligible, and the line as one of zero dissipation. These methods will be studied separately.

STANDING WAVES

When the transmission line is not matched with its load i.e., load impedance is not equal to the characteristic impedance ($Z_R = Z_0$), the energy delivered to the load is reflected back to the source.

The combination of incident and reflected waves give rise to the standing waves.

STANDING-WAVE RATIO

The measurement of standing waves on a transmission line yields information about equipment operating conditions. Maximum power is absorbed by the load when $Z_L = Z_0$. If a line has no standing waves, the termination for that line is correct and maximum power transfer takes place.

$$VSWR = \frac{|V_{MAX}|}{|V_{MIN}|}$$

You have probably noticed that the variation of standing waves shows how near the rf line is to being terminated in Z_0 . A wide variation in voltage along

the length means a termination far from Z_0 . A small variation means termination near Z_0 . Therefore, the ratio of the maximum to the minimum is a measure of the perfection of the termination of a line. This ratio is called the STANDING-WAVE RATIO (SWR) and is always expressed in whole numbers. For example, a ratio of 1:1 describes a line terminated in its characteristic impedance (Z_0).

Voltage Standing-Wave Ratio

The ratio of maximum voltage to minimum voltage on a line is called the VOLTAGE STANDING-WAVE RATIO (VSWR). Therefore: The vertical lines in the formula indicate that the enclosed quantities are absolute and that the two values are taken without regard to polarity, Depending on the nature of the standing waves, the numerical value of VSWR ranges from a value of 1 ($Z_L = Z_0$, no standing waves) to an infinite value for theoretically complete reflection.

Since there is always a small loss on a line, the minimum voltage is never zero and the VSWR is always some finite value. However, if the VSWR is to be a useful quantity, the power losses along the line must be small in comparison to the transmitted power voltage. Since power is proportional to the square of the voltage, the ratio of the square of the maximum and minimum voltages is called the power standing-wave ratio. In a sense, the name is misleading because the power along a transmission line does not vary.

Current Standing-Wave Ratio

The ratio of maximum to minimum current along a transmission line is called CURRENT STANDING-WAVE RATIO (ISWR). Therefore: This ratio is the same as that for voltages. It can be used where measurements are made with loops that sample the magnetic field along a line. It gives the same results as VSWR measurements.

STANDING WAVE RATIO

The ratio of the maximum to minimum magnitudes of voltage or current on a line having standing waves is called the standing wave ratio or voltage standing wave ratio (VSWR)

$$S = \frac{|V_{\max}|}{|V_{\min}|} = \frac{|I_{\max}|}{|I_{\min}|}$$

Voltage equation is

$$V = \frac{V_R (Z_R + Z_0)}{2Z_R} [e^{j\beta z} + K e^{-j\beta z}]$$

Maxima of voltage occurs at which the incident and reflected waves are in phase

$$V_{\max} = \frac{V_R (Z_R + Z_0)}{2Z_R} [1 + K]$$

$\frac{\lambda}{4}$

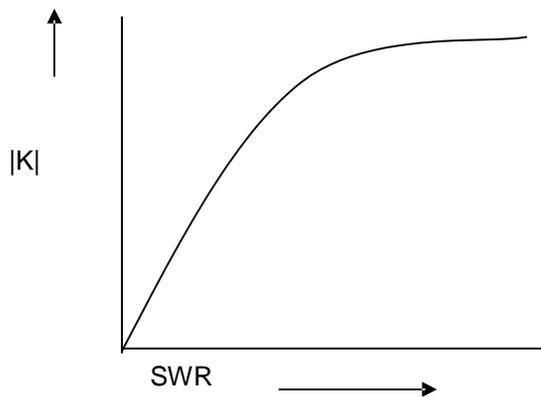
Minima of voltage occurs at which the incident and reflected waves are out of phase

$$V_{\min} = \frac{V_R(Z_R + Z_0)}{2Z_R} [1 - K]$$

$$\frac{V_{\max}}{V_{\min}} = \frac{1 + |K|}{1 - |K|}$$

$$K = \left| \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}} \right|$$

$$K = \left| \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}} \right|$$



This figure shows the relation between standing wave ratio S and reflection coefficient

ONE EIGHTH WAVELENGTH

For the transmission line the voltage and current at any point x from the receiving end of the transmission line is

$$V = \frac{V_R(Z_R + Z_0)}{2Z_R} [e^{\gamma x} + Ke^{-\gamma x}]$$

$$I = \frac{I_R(Z_R + Z_0)}{2Z_0} [e^{\gamma x} - Ke^{-\gamma x}]$$

The term with γx is identified as the incident wave progressing forward from the source to the load, whereas the term involving $e^{-\gamma x}$ is the reflected wave traveling from load back towards the source.

For the line of zero dissipation, the attenuation constant is zero.

$$\gamma = j\beta \quad \text{and } Z_0 = R_0$$

$$V = \frac{V_R (Z_R + Z_0)}{2Z_R} [e^{j\beta x} + K e^{-j\beta x}]$$

After simplification of the above equation for standing wave $|K| = 1$

$$V = V_R \cos \beta x + j I_R R_0 \sin \beta x.$$

Similarly, for the current on the transmission line

$$I = I_R \cos \beta x + j V_R / R_0 \sin \beta x.$$

The input impedance of a dissipation line is

$$Z_s = \frac{V}{I}$$

$$= \frac{V_R \cos \beta x + j I_R R_0 \sin \beta x}{I_R \cos \beta x + j V_R / R_0 \sin \beta x}$$

$$= R_0 \frac{Z_R \cos \beta x + j R_0 \sin \beta x}{R_0 \cos \beta x + j Z_R \sin \beta x}$$

$$Z = \frac{Z_R + j R_0 \tan \beta x}{R_0 + j Z_R \tan \beta x}$$

Or

For an eighth wave line

$$x = \lambda / 8, \beta x = \frac{2\pi}{\lambda} \cdot \frac{\lambda}{8}$$

$$Z_s = \frac{Z_R + j R_0 \tan(\pi / 4)}{R_0 + j Z_R \tan(\pi / 4)}$$

$$Z_s = \frac{Z_R + j R_0}{R_0 + j Z_R}$$

If such a line is terminated with pure resistance

$$Z_R = R_R$$

$$Z_s = \frac{R_R + j R_0}{R_0 + j R_R}$$

Since, both the numerator and denominator have identical magnitudes, then

$$|Z_s| = R_0$$

Thus an eighth – wave line may be used to transfer any resistance to impedance with a magnitude equal to R_0 of the line, or obtain a magnitude match between a resistance of any value and R_0 , the internal resistance of the source

QUARTER WAVELINE AND IMPEDANCE MATCHING

The input impedance of a dissipation transmission line is

$$Z_s = \frac{Z_R + jR_0 \tan \beta x}{R_0 + jZ_R \tan \beta x}$$

$$Z_s = \frac{Z_R / \tan \beta x + jR_0}{R_0 / \tan \beta x + jZ_R}$$

For a quarter wave line

$$x = \lambda / 4, \beta x = 2\pi / \lambda * \lambda / 4 = \lambda / 2$$

Substituting the parameter value in the above equation the sending end impedance of the quarter wave transformer is

$$Z_s = \frac{R_0^2}{Z_R}$$

A quarter wave section of line may be considered as a transformer to match a load of Z_R to a source of Z_s . Such a match can be obtained if the characteristic impedance R_0 of the matching quarter wave section of the line is properly chosen.

$$R_0 = \sqrt{Z_s Z_R}$$

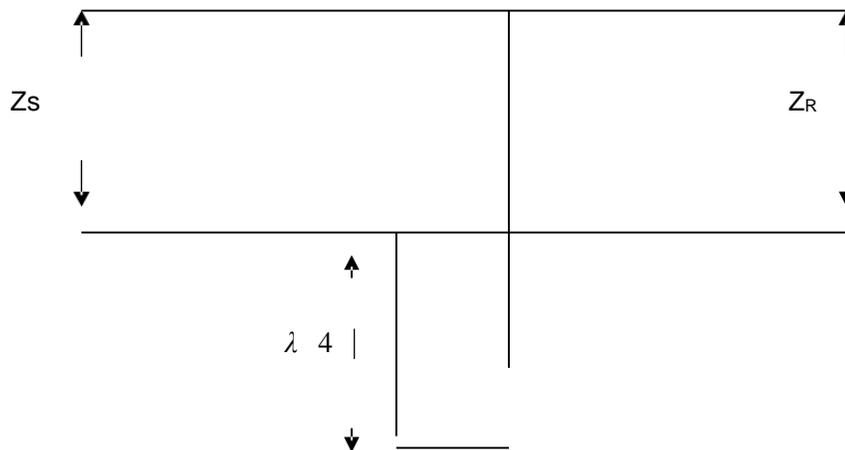
UNIT III

IMPEDANCE MATCHING IN HIGH FREQUENCY

APPLICATIONS OF QUARTER WAVE TRANSFORMER

- A quarter wave transformer may also be used if the load is not a pure resistance
- The quarter wave transformer is a single frequency or narrow band device. The bandwidth may be increased by using two or more Quarter wave section in series.
- A quarter wave transformer may be considered as an impedance inverter in that it can transform a low impedance into a high impedance and vice versa.

35



HALF WAVE LINE

Already we know that The input impedance of a dissipation less line is

$$Z_s = \frac{Z_R \cos \beta x + jR_0 \tan \beta x}{R_0 \cos \beta x + jZ_R \tan \beta x}$$

For a quarter wave line

$$x = \lambda / 2, \beta x = 2\pi / \lambda * \lambda / 2 = \pi$$

$$Z_s = \frac{Z_R \cos \pi + jR_0 \tan \pi}{R_0 \cos \pi + jZ_R \tan \pi}$$

$$R_0 + jZ_R$$

$$R_0 R_0$$

$$Z_R$$

$\tan \pi$ 

www.padeepz.net

$$Z_s = Z_R$$

This line may be considered as one to one transformer.

APPLICATION:

It is used in connecting a load to a source in cases when the load and source cannot be made adjacent.

STUB MATCHING

Need for stub matching

To match the load impedance to be equal to the input impedance.

TYPES

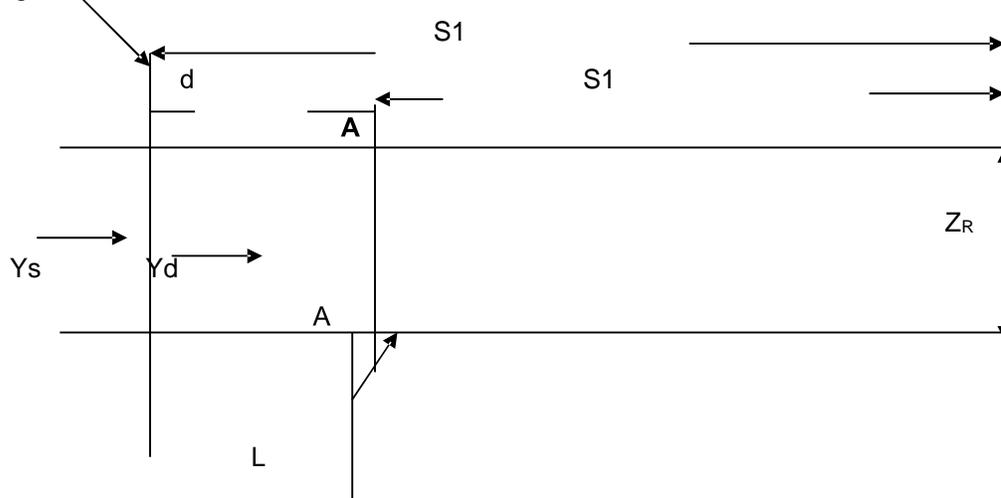
1. SINGLE STUB MATCHING
2. DOUBLE STUB MATCHING

SINGLE STUB MATCHING ON A LINE

For greatest efficiency and delivered power, a high frequency transmission line should be operated as a smooth line or with an R_0 termination. However, the usual loads, such as antennas, do not in general have resistance of value equal to R_0 , so that many cases it is necessary to introduce some form of impedance – transforming action between line and load to make the load appear to the line as a resistance value R_0 .

The quarter wave line or transformer and the tapered line are such impedance matching devices. Another means of accomplishing the desired result is the use of an open or closed stub line of suitable length as a reactance shunted across the transmission line at the load to resonance with an antiresonant resistance equal to R_0 .

Voltage minimum before insertion of the stub



Since the input conductance of the of a line is S/R_0 at a voltage maximum and S/R_0 at a voltage minimum, then at some intermediate [point A the real part of the input admittance may be an intermediate value of $1/R_0$ or the input admittance at A has a value

$$Y_s = \frac{1}{R_0} \pm jB$$

The susceptance B is the shunt value at the point in question. After the point having a conductance equal to $1/R_0$ is located, a short stub line having input susceptance of $\pm B$ may be connected across the transmission line. The input admittance at this point then is

$$Y_s = \frac{1}{R_0} \pm jB \pm jB = \frac{1}{R_0}$$

Or the input impedance of the line at point A looking towards the load is

$$Z_s = R_0$$

37

Since both the location and length of the stub must be determined, two independent measurements must be made on the original line and load to secure sufficient data.

The most easily obtained measurements are the standing wave ratio S and the position of a voltage minimum, usually the minimum nearest to the load. A voltage minimum is chosen rather than a maximum, since its position usually can be determined more accurately.

If the location of the stub is fixed with respect to an original voltage minimum, no knowledge of the load impedance is needed.

Because of the paralleling of elements, it is most convenient to work with admittances. The input impedance equation is looking towards the load from any point on the line, may written as

$$Y_s = \frac{1}{R_0} \frac{1 - jK \tan(\varphi - 2\beta s)}{1 + jK \tan(\varphi - 2\beta s)}$$

Writing $G_0=1/R_0$ and changing to rectangular coordinates gives

$$Y_s = G_0 \frac{1 - jK \cos(\varphi - 2\beta s) - jK \sin(\varphi - 2\beta s)}{1 + jK \cos(\varphi - 2\beta s) + jK \sin(\varphi - 2\beta s)}$$

And upon rationalizing,

$$Y_s = G_0 \frac{1 - K^2 - 2jK \sin(\varphi - 2\beta s)}{1 + K^2 + 2jK \cos(\varphi - 2\beta s)}$$

Expressing the shunt conductance as a dimensionless ratio G_s/G_0 , or on a per unit basis,

$$\frac{G_s}{G_0} = \frac{1 - |K|^2}{1 - 2|K| \cos(\varphi - 2\beta s) + |K|^2}$$

And the shunt susceptance on a per unit basis is

$$B_s = G_0 \frac{1 - |K|^2}{1 - 2|K| \cos(\varphi - 2\beta s) + |K|^2}$$

$$\frac{1 - |K|^2 - 2j|K| \sin(\varphi - 2\beta s)}{|K|^2 + 2j|K| \cos(\varphi - 2\beta s)}$$

After simplifying the above equations we get the location and distance of the stub

The distance d from the voltage minimum to the point of stub connection is

$$d = s_2 - s_1$$

$$d = \frac{\cos^{-1}|K|}{2\beta}$$

$$= \frac{s_1 + \frac{\lambda}{4}}{\pi}$$

Before connection of the stub the equation is

$$L = \frac{\lambda}{2\pi} \tan^{-1} \frac{S\sqrt{}}{S-1}$$

38

The stub length should be

$$L' = \frac{\lambda}{2} - L$$

The Circle diagram for dissipation less line

A somewhat similar circle diagram may be obtained, however, that solves the impedance equation and simplifies the design of dissipationless lines considerably. The input impedance equation for a dissipation less line may be written as

$$\frac{Z_s}{R_0} = \frac{1 + |K| \cos(\varphi - 2\beta s)}{1 - |K| \cos(\varphi - 2\beta s)}$$

$$\frac{Z_s}{R_0} = r_a + jx_a$$

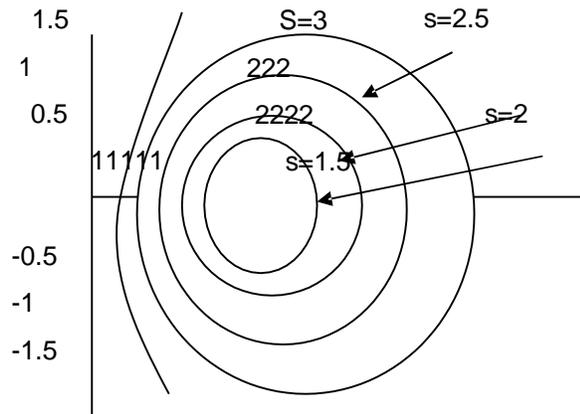
An actual circle will have the radius

$$r = \frac{S - 1}{2S} = \frac{S - 1}{2}$$

And the center of the circle on the positive is

$$c = \frac{S + 1}{2S} = \frac{S + 1}{2}$$

A family of circles may be drawn for successive values of S as in fig. In drawing particular circles it is interesting to note that for any circle the intercept near origin is at 1/S, and that far removed from the origin is at S units on the ra axis



The minimum value for S is unity. The above figure shows that all S circles must surround the 1,0 point. In fact, the circle for S = 1 is represented by the 1,0 point.

39

The maximum value of S is infinity, for the case of open circuit or short circuit line termination. As S increase, the radius of the S increases, and the center moves to the right; for the limiting case of S = infinity, the circle becomes the x_a axis.

The line impedance is maximum. And

$$\frac{Z_s}{R_0} = \frac{1 + |K|}{1 - |K|}$$

When $\frac{Z_s}{R_0}$ terminates at the circle intercept 1/S, the line impedance has a minimum

value, and

$$\frac{Z_s}{R_0 S} = \frac{1 - |K|}{1 + |K|}$$

After some simplification we get the final equation, that

$$\frac{r_a + jx_a}{R_0} = 1 + \frac{1}{\tan^2 \beta s}$$

$$= \frac{1}{\sin^2 2\beta s}$$

Lines of equal βs are the seen to be circles of radius

$$= \frac{1}{\sin^2 2\beta s}$$

With the shift of center downward on the xa axis (ordinate)

$$= - \frac{1}{\tan^2 2\beta s}$$

S- CIRCLE

$$\frac{g_a - S^2}{1 + S^2} + b_a = \frac{S^2 + 1}{2S}$$

Application of the circle diagram

- Used to find the input impedance of a line of any chosen length

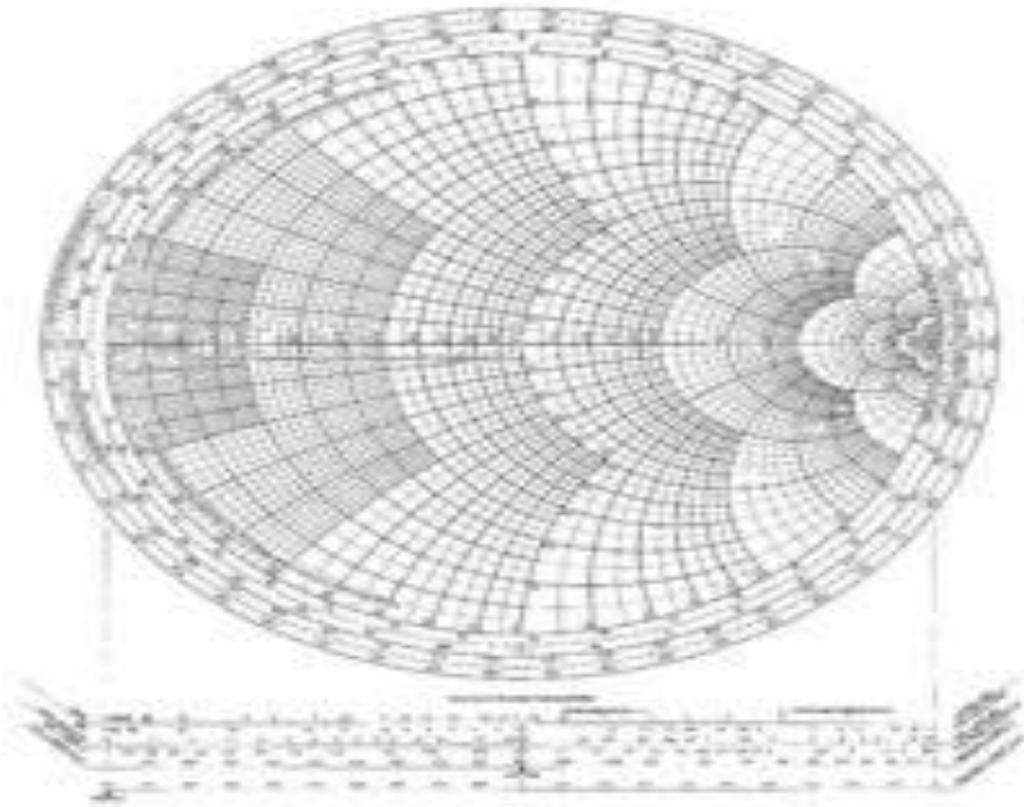
An open circuited line has $S = \infty$, the corresponding S circle appearing as the vertical axis. The input impedance is then pure reactance, with the value for various electrical lengths determined by the intersections of the corresponding βs with the vertical axis.

- The input admittance of the line may be found by this method.

A short circuited line may be solved by determining its admittance. The S circle is again the vertical axis, and susceptance values may be read off at appropriate intersections of the βs circles with the vertical axis.

SMITH CHART

The Complete Smith Chart
Black Magic Design



An impedance Smith chart (with no data plotted)

The Smith Chart can be used to represent many parameters including impedances, admittances, reflection coefficients, scattering parameters, noise figure circles, constant gain contours and regions for unconditional stability. The Smith Chart is most frequently used at or within the unity radius region. However, the remainder is still mathematically relevant, being used, for example, in oscillator design and stability analysis

The Smith Chart is plotted on the complex reflection coefficient plane in two dimensions and is scaled in normalized impedance (the most common), normalized admittance or both, using different colors to distinguish between them. These are often known as the Z, Y and YZ Smith Charts respectively.^[7] Normalized scaling allows the Smith Chart to be used for problems involving any characteristic impedance or system impedance, although by far the most commonly used is 50 ohms. With relatively simple graphical construction it is straightforward to convert between normalized impedance (or normalized admittance) and the corresponding complex voltage reflection coefficient.

The Smith Chart has circumferential scaling in wavelengths and degrees. The wavelengths scale is used in distributed component problems and represents the distance measured along the transmission line connected between the generator or source and the load to the point under consideration. The degrees scale represents the angle of the voltage reflection coefficient at that point. The Smith Chart may also be used for lumped element matching and analysis problems.

Use of the Smith Chart and the interpretation of the results obtained using it requires a good understanding of AC circuit theory and transmission line theory, both of which are prerequisites for RF engineers.

As impedances and admittances change with frequency, problems using the Smith Chart can only be solved manually using one frequency at a time, the result being represented by a point. This is often adequate for narrow band applications (typically up to about 5% to 10% bandwidth) but for wider bandwidths it is usually necessary to apply Smith Chart techniques at

more than one frequency across the operating frequency band. Provided the frequencies are sufficiently close, the resulting Smith Chart points may be joined by straight lines to create a locus.

A locus of points on a Smith Chart covering a range of frequencies can be used to visually represent:

- How capacitive or how inductive a load is across the frequency range

- How difficult matching is likely to be at various frequencies
- How well matched a particular component is.

The accuracy of the Smith Chart is reduced for problems involving a large spread of impedances or admittances, although the scaling can be magnified for individual areas to accommodate these.

Regions of the Z Smith Chart

If a polar diagram is mapped on to a cartesian coordinate system it is conventional to measure angles relative to the positive x-axis using a counter-clockwise direction for positive angles. The magnitude of a complex number is the length of a straight line drawn from the origin to the point representing it.

The Smith Chart uses the same convention, noting that, in the normalized impedance plane, the positive x-axis extends from the center of the Smith Chart at to the point . The region above the x-axis represents inductive impedances and the region below the x-axis represents capacitive impedances. Inductive impedances have positive imaginary parts and capacitive impedances have negative imaginary parts.

If the termination is perfectly matched, the reflection coefficient will be zero, represented effectively by a circle of zero radius or in fact a point at the centre of the Smith Chart. If the termination was a perfect open circuit or short circuit the magnitude of the reflection coefficient would be unity, all power would be reflected and the point would lie at some point on the unity circumference circle.

Circles of Constant Normalized Resistance and Constant Normalized Reactance

The normalized impedance Smith Chart is composed of two families of circles: circles of constant normalized resistance and circles of constant normalized reactance. In the complex reflection coefficient plane the Smith Chart occupies a circle of unity radius centered at the origin. In Cartesian coordinates therefore the circle would pass through the points (1,0) and (-1,0) on the x-axis and the points (0,1) and (0,-1) on the y-axis.

Working with both the Z Smith Chart and the Y Smith Charts

In RF circuit and matching problems sometimes it is more convenient to work with admittances (representing conductance's and susceptances) and sometimes it is more

convenient to work with impedances (representing resistances and reactance's). Solving a

typical matching problem will often require several changes between both types of Smith Chart, using normalized impedance for series elements and normalized admittances for parallel

elements. For these a dual (normalized) impedance and admittance Smith Chart may be used. Alternatively, one type may be used and the scaling converted to the other when required.

In order to change from normalized impedance to normalized admittance or vice versa, the point representing the value of reflection coefficient under consideration is moved through

43

exactly 180 degrees at the same radius. For example the point P1 in the example representing a reflection coefficient of has a normalized impedance of. To graphically change this to the equivalent normalized admittance point, say Q1, a line is drawn with a ruler from P1 through the Smith Chart centre to Q1, an equal radius in the opposite direction. This is equivalent to moving the point through a circular path of exactly 180 degrees. Reading the value from the Smith Chart for Q1, remembering that the scaling is now in normalized admittance, gives .

Once a transformation from impedance to admittance has been performed the scaling changes to normalized admittance until such time that a later transformation back to normalized impedance is performed.

PROBLEMS ON SINGLE STUB MATCHING

1. Determine the length and the distance of the stub from the load. Given that a complex load $Z_L = 50 - j100$ is to be matched to a 75 ohm transmission line using a short circuited stub.

Given

Characteristic impedance of the transmission line $Z_0 = 75\text{ohm}$

Load impedance to be matched to the transmission line $Z_L = 50 - j100$

To find

1. Distance of the stub from the load

2. Length of the stub from the load

Solution

1. The normalized impedance is determined by dividing the load impedance by the characteristic impedance of the transmission line.

$$Z_L = \frac{Z_L}{Z_0} = \frac{50 - j100}{75} = 0.667 - j1.33$$

2. The normalized impedance, Z_L is plotted on the smith chart by determining the point of intersection between the constant R circle with $R = 0.667$ and constant X circle with $X = 1.33$

The impedance circle is drawn.

Because the stubs are connected in parallel with the load, admittances can be much easily used rather than impedances to simplify the calculations.

3. The normalized admittance is determined from the smith chart by simply rotating the impedance plot, by 180 degree. This is simply done by drawing a line from point A

to the opposite side of the circle, point B.

4. the admittance point is rotated clockwise to a point on the impedance circle where it intersects the characteristic impedance Z_0 . At the point C. The real component of the input impedance at this point is equal to the characteristic impedance Z_0 . At this point C, the admittance is $y=1+j1.7$.
5. The distance from point B to point C, in terms of the wavelength is how far from load the stub must be placed,
The stub must have a zero resistive component impedance and susceptance that has the opposite polarity.
6. To determine the length of the shorted stub that has an opposite reactive component to the input admittance, the outside of the Smith chart ($R=0$) is moved around with the starting point at D {since at point D $t=0$ and hence $\gamma = \infty$ }, until an admittance $y = 1.7j$ is found
7. The distance between point D and E is the length of the stub. For this quantity the from the smith chart,

DOUBLE STUB MATCHING

2. Using Double stub matching, match a complex load of $Z_L = 18.75+j56.25$ to a line with characteristic impedance $Z_0 = 75\text{ohm}$.

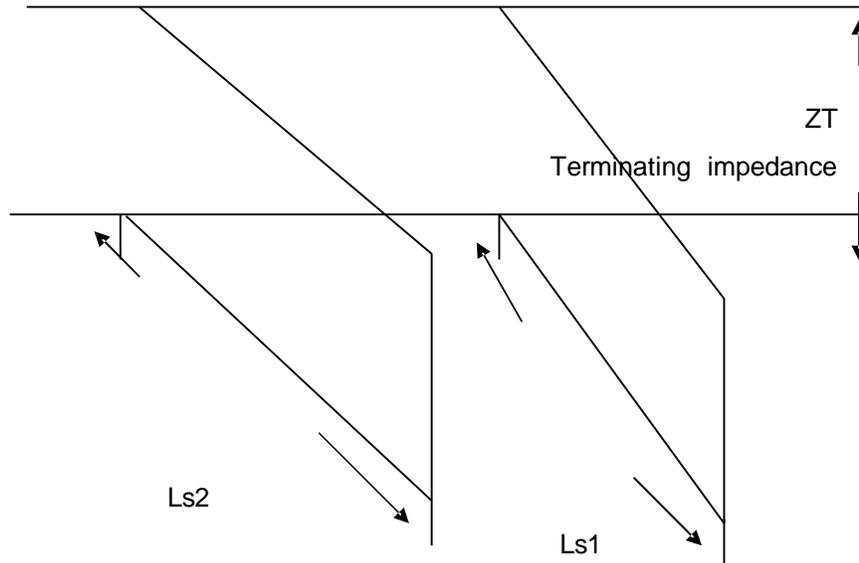
Determine the stub lengths, assuming a quarter wavelength spacing are maintained between the two short circuited stubs.

A spacing of $\lambda / 4$ is maintained between the stubs, stub2 and stub1. For smooth line operation of the transmission line the input impedance looking into the terminals 2,2 of the line should be,

$$Y_{2,2} = 1/ Z_0$$

The stub at 1,1 must be capable to transform the admittance at the terminating impedance end to the circle B which is displaced from the circle A; $R=1$ by ' $\lambda / 4$ '.

The quarter wavelength line will further transform the admittance into a value at 2,2 which plot on the circle A. Thus the line to load distance between position 2,2 is not required to be determined.



The normalized load impedance

$$Z_L = \frac{Z_i}{Z_0} = \frac{18.75 + j56.25}{75}$$

$$Z_L = 0.25 + j0.75$$

Plotting the normalized impedance on the Smith chart, the impedance circle is drawn with distance between the point (1,0) and the point of the normalized impedance as the radius {distance, OA}

1. Moving by 180 degree (0.25λ) on the impedance circle, that is at a diametrically opposite point to the point A, i.e., point B will give the normalized admittance. From the smith chart $Y_L = 0.4 - j1.2$
2. Circle A is the constant R circle for $R = 1$. Circle B is the locus of all the points on the circle A is displaced by $\lambda/4$, quarter wavelength. The stub 1 adds a susceptance of all the points on the circle B. Since stub 1 cannot alter the conductance, to a point on the circle B, point C,

$$Y(\text{at point C}) = 0.4 - j0.5$$

3. Transferring the point C to the point D on the circle A, since the line between 1,1 and 2,2 is a quarter wave line that transforms the admittance at 1,1 to 2,2 such that the conductance equals the characteristic conductance, $1/Z_0$.

$$Y(\text{At point D}) = 1.0 + j1.2$$

4. The stub length at 2,2 should cancel the imaginary part of the above admittance of the stub at 2,2 must be -1.2 .

5. To find the length of the stub with an admittance ,

$$(a) +j0.7 \text{ and } (b) -j1.2$$

The outside circle of the smith chart (the circle, $R=0$), is moved around having a reference at a point P, until

An admittance $y = -1.2$ is found at point E and

An admittance $y = +0.7$ is found at point F.

6. From the smith chart,

Length of the stub 1 = distance between P and F $L_{s1} = 0.348 \lambda$

Length of the stub 2 = distance between P and E $L_{s2} = 0.11 \lambda$

3. Determine the following:

- (a) Standing wave ratio (VSWR)
 (b) Load Admittance
 (c) Impedance of the transmission line at the maximum and minimum of the stationary waves along the line
 (d) Distance between load and first voltage maximum. For a transmission line with characteristic impedance of 50 ohm with a receiving end of $100 + j121$. The wavelength of the electrical signal along the line is 2.5m.

Given:

Characteristic impedance $Z_0 = 50$ ohm

Load impedance $Z_L = 100 + j121$ ohm

Wavelength of the electrical signal $\lambda = 2.5$

Solution

1. Normalized impedance = $\frac{100 + j121}{50} = 2 + j2.42$

Plotting the point p on the smith chart . The impedance circle is drawn with O(1+j0) centre and radius as (OP), the distance between centre and the normalized voltage standing Wave Ratio = 5

2. The point Q diametrically opposite to the normalized impedance point on the impedance circle is the normalized admittance of the load.

$$\begin{aligned} \frac{Y}{G_0} &= 0.22 - j0.25 \\ \text{Load impedance } YZ_0 &= 0.22 - j0.25 \\ Y &= \frac{0.22 - j0.25}{50} \\ &= 0.0044 - j0.005 \text{ mho} \end{aligned}$$

3. Impedance at the first voltage maximum from load =

$$\begin{aligned} &5 \times Z_0 \\ &= 250 \text{ ohm} \end{aligned}$$

Impedance at the first voltage minimum=

$$\begin{aligned} &0.2 \times Z_0 \\ &= 10 \text{ ohm} \end{aligned}$$

Distance between load and first voltage maximum=

$$\begin{aligned} &0.042\lambda \\ &= 0.042 \times 2.5 \text{ m} \\ &= 0.105 \text{ m} \end{aligned}$$

UNIT -IV PASSIVE FILTERS

1. Neper

A neper (Symbol: Np) is a logarithmic unit of ratio. It is not an SI unit but is accepted for use alongside the SI. It is used to express ratios, such as gain and loss, and relative values. The name is derived from John Napier, the inventor of logarithms.

Like the decibel, it is a unit in a logarithmic scale, the difference being that where the decibel uses base-10 logarithms to compute ratios, the neper uses base $e \approx 2.71828$. The value of a ratio in nepers, Np, is given by

$$L_{Np} = \ln \frac{x_1}{x_2} = \ln x_1 - \ln x_2.$$

where x_1 and x_2 are the values of interest, and \ln is the natural logarithm.

The neper is often used to express ratios of voltage and current amplitudes in electrical circuits (or pressure in acoustics), whereas the decibel is used to express power ratios. One kind of ratio may be converted into the other. Considering that wave power is proportional to the square of the amplitude, we have

$$1 \text{ Np} = \frac{20}{\ln 10} \text{ dB} = 20 \log_{10} e \text{ dB} \approx 8.685889638 \text{ dB}$$

and

$$1 \text{ dB} = \frac{\ln 10}{20} \text{ Np} = \frac{1}{20 \log_{10} e} \text{ Np} \approx 0.115129254 \text{ Np}.$$

The decibel and the neper have a fixed ratio to each other. The (voltage) level is

$$\begin{aligned} L &= 10 \lg \frac{x_1^2}{x_2^2} && \text{dB} \\ &= 10 \lg \left(\frac{x_1}{x_2} \right)^2 && \text{dB} \\ &= 20 \lg \frac{x_1}{x_2} && \text{dB} \\ &= \ln \frac{x_1}{x_2} && \text{Np}. \end{aligned}$$

Like the decibel, the neper is a dimensionless unit. The ITU recognizes both units.

2. Decibel

The decibel (dB) is a logarithmic unit of measurement that expresses the magnitude of a physical quantity (usually power or intensity) relative to a specified or implied reference level. Since it expresses a ratio of two quantities with the same unit, it is a dimensionless unit. A decibel is one tenth of a bel, a seldom-used unit.

The decibel is widely known as a measure of sound pressure level, but is also used for a wide variety of other measurements in science and engineering (particularly acoustics, electronics, and control theory) and other disciplines. It confers a number of advantages, such as the ability to conveniently represent very large or small numbers, a logarithmic scaling that roughly corresponds to the human perception of sound and light, and the ability to carry out multiplication of ratios by simple addition and subtraction.

The decibel symbol is often qualified with a suffix, which indicates which reference quantity or frequency weighting function has been used. For example,

"dBm" indicates that the reference quantity is one milliwatt, while "dBu" is referenced to 0.775 volts RMS.^[1]

The definitions of the decibel and bel use base-10 logarithms. For a similar unit using natural logarithms to base e, see neper.

Definitions

A decibel is one-tenth of a bel, i.e. 1 B=10 dB. The bel (B) is the logarithm of the ratio of two power quantities of 10:1, and for two field quantities in the ratio $\sqrt{10} : 1$

^[8]. A field quantity is a quantity such as voltage, current, sound pressure, electric field strength, velocity and charge density, the square of which in linear systems is proportional to power. A power quantity is a power or a quantity directly proportional to power, e.g. energy density, acoustic intensity and luminous intensity.

The calculation of the ratio in decibels varies depending on whether the quantity being measured is a power quantity or a field quantity.

Power quantities

When referring to measurements of power or intensity, a ratio can be expressed in decibels by evaluating ten times the base-10 logarithm of the ratio of the measured quantity to the reference level. Thus, if L represents the ratio of a power value P_1 to another power value P_0 , then L_{dB} represents that ratio expressed in decibels and is calculated using the formula:

$$L_{dB} = 10 \log_{10} \left(\frac{P_1}{P_0} \right)$$

P_1 and P_0 must have the same dimension, i.e. they must measure the same type of quantity, and the same units before calculating the ratio: however, the choice of scale for this common unit is irrelevant, as it changes both quantities by the same factor, and thus cancels in the ratio—the ratio of two quantities is scale-invariant. Note that if $P_1 = P_0$ in the above equation, then $L_{dB} = 0$. If P_1 is greater than P_0 then L_{dB} is positive; if P_1 is less than P_0 then L_{dB} is negative.

Rearranging the above equation gives the following formula for P_1 in terms of P_0

and L_{dB} :

$$P_1 = 10^{\frac{L_{dB}}{10}} P_0.$$

Since a bel is equal to ten decibels, the corresponding formulae for measurement in bels (L_B) are

$$L_B = \log_{10} \left(\frac{P_1}{P_0} \right)$$

$$P_1 = 10^{L_B} P_0.$$

Field quantities

When referring to measurements of field amplitude it is usual to consider the ratio of the squares of A_1 (measured amplitude) and A_0 (reference amplitude). This is because in most applications power is proportional to the square of amplitude, and it is desirable for the two decibel formulations to give the same result in such typical cases. Thus the following definition is used:

$$L_{dB} = 10 \log_{10} \left(\frac{A_1^2}{A_0^2} \right) = 20 \log_{10} \left(\frac{A_1}{A_0} \right).$$

This formula is sometimes called the 20 log rule, and similarly the formula for ratios of powers is the 10 log rule, and similarly for other factors.^[citation needed] The equivalence of $10 \log_{10} \frac{a^2}{b^2}$ and $20 \log_{10} \frac{a}{b}$ is of the standard properties of logarithms.

The formula may be rearranged to give

$$A_1 = 10^{\frac{L_{dB}}{20}} A_0$$

Similarly, in electrical circuits, dissipated power is typically proportional to the square of voltage or current when the impedance is held constant. Taking voltage as an example, this leads to the equation:

$$G_{dB} = 20 \log_{10} \left(\frac{V_1}{V_0} \right)$$

where V_1 is the voltage being measured, V_0 is a specified reference voltage, and G_{dB} is the power gain expressed in decibels. A similar formula holds for current.

10 Log X	X
100	10000000000
90	1000000000
80	100000000
70	10000000
60	1000000
50	100000
40	10000
30	1000
20	100
10	10
0	1
-10	0.1
-20	0.01
-30	0.001
-40	0.0001
-50	0.00001
-60	0.000001
-70	0.0000001
-80	0.00000001
-90	0.000000001
-100	0.0000000001



An example scale showing x and $10 \log x$. It is easier to grasp and compare 2 or 3 digit numbers than to compare up to 10 digits.

Note that all of these examples yield dimensionless answers in dB because they are relative ratios expressed in decibels.

- To calculate the ratio of 1 kW (one kilowatt, or 1000 watts) to 1 W in decibels, use the formula

$$G_{dB} = 10 \log_{10} \left(\frac{1000 \text{ W}}{1 \text{ W}} \right) = 30 \text{ dB}$$

- To calculate the ratio of $\sqrt{(1000)} \text{ V} \approx 31.62 \text{ V}$ to 1 V in decibels, use the formula

$$G_{\text{dB}} = 20 \log_{10} \left(\frac{31.62 \text{ V}}{1 \text{ V}} \right) = 30 \text{ dB}$$

Notice that $(31.62 \text{ V}/1 \text{ V})^2 \approx 1 \text{ kW}/1 \text{ W}$, illustrating the consequence from the definitions above that G_{dB} has the same value, **30 dB**, regardless of whether it is obtained with the 10-log or 20-log rules; provided that in the specific system being considered power ratios are equal to amplitude ratios squared.

- To calculate the ratio of 1 mW (one milliwatt) to 10 W in decibels, use the formula

$$G_{\text{dB}} = 10 \log_{10} \left(\frac{0.001 \text{ W}}{10 \text{ W}} \right) = -40 \text{ dB}$$

- To find the power ratio corresponding to a 3 dB change in level, use the formula

$$G = 10^{\frac{3}{10}} \times 1 = 1.99526... \approx 2$$

A change in power ratio by a factor of 10 is a 10 dB change. A change in power ratio by a factor of two is approximately a 3 dB change. More precisely, the factor is $10^{3/10}$, or 1.9953, about 0.24% different from exactly 2. Similarly, an increase of 3 dB implies an increase in voltage by a factor of approximately $\sqrt{2}$ or about 1.41, an increase of 6 dB corresponds to approximately four times the power and twice the voltage, and so on. In exact terms the power ratio is $10^{6/10}$, or about 3.9811, a relative error of about 0.5%.

Merits

The use of the decibel has a number of merits:

- The decibel's logarithmic nature means that a very large range of ratios can be represented by a convenient number, in a similar manner to scientific notation. This allows one to clearly visualize huge changes of some quantity. (See Bode Plot and half logarithm graph.)

- The mathematical properties of logarithms mean that the overall decibel gain of a multi-component system (such as consecutive amplifiers) can be calculated simply by summing the decibel gains of the individual components, rather than needing to multiply amplification factors. Essentially this is because $\log(A \times B \times C \times \dots) = \log(A) + \log(B) + \log(C) + \dots$
- The human perception of, for example, sound or light, is, roughly speaking, such that a doubling of actual intensity causes perceived intensity to always increase by the same amount, irrespective of the original level. The decibel's logarithmic scale, in which a doubling of power or intensity always causes an increase of approximately 3 dB, corresponds to this perception.

Absolute and relative decibel measurements

Although decibel measurements are always relative to a reference level, if the numerical value of that reference is explicitly and exactly stated, then the decibel measurement is called an "absolute" measurement, in the sense that the exact value of the measured quantity can be recovered using the formula given earlier. For example, since dBm indicates power measurement relative to 1 milliwatt,

- 0 dBm means no change from 1 mW. Thus, 0 dBm is the power level corresponding to a power of exactly 1 mW.
- 3 dBm means 3 dB greater than 0 dBm. Thus, 3 dBm is the power level corresponding to $10^{3/10} \times 1$ mW, or approximately 2 mW.
- -6 dBm means 6 dB less than 0 dBm. Thus, -6 dBm is the power level corresponding to $10^{-6/10} \times 1$ mW, or approximately 250 μ W (0.25 mW).

If the numerical value of the reference is not explicitly stated, as in the dB gain of an amplifier, then the decibel measurement is purely relative. The practice of attaching a suffix to the basic dB unit, forming compound units such as dBm, dBu, dBA, etc, is not permitted by SI.^[10] However, outside of documents adhering to SI units, the practice is very common as illustrated by the following examples.

Absolute measurements

Electric power

dBm or dBmW

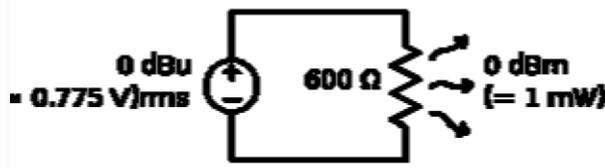
dB(1 mW) — power measurement relative to 1 milliwatt. $X_{\text{dBm}} = X_{\text{dBW}} + 30$.

dBW

dB(1 W) — similar to dBm, except the reference level is 1 watt. 0 dBW = +30 dBm; -30 dBW = 0 dBm; $X_{\text{dBW}} = X_{\text{dBm}} - 30$.

Voltage

Since the decibel is defined with respect to power, not amplitude, conversions of voltage ratios to decibels must square the amplitude, as discussed above.



A schematic showing the relationship between dBu (the voltage source) and dBm (the power dissipated as heat by the 600 Ω resistor)

dBV

dB(1 V_{RMS}) — voltage relative to 1 volt, regardless of impedance.^[1]

dBu or dBv

dB(0.775 V_{RMS}) — voltage relative to 0.775 volts.^[1] Originally dBv, it was changed to dBu to avoid confusion with dBV.^[11] The "v" comes from "volt", while "u" comes from "unloaded". dBu can be used regardless of impedance, but is derived from a 600 Ω load dissipating 0 dBm (1 mW). Reference

$$\text{voltage } V = \sqrt{600 \Omega \cdot 0.001 \text{ W}} \approx 0.7746 \text{ V}$$

dBmV

dB(1 mV_{RMS}) — voltage relative to 1 millivolt across 75 Ω ^[12]. Widely used in cable television networks, where the nominal strength of a single TV signal at the receiver terminals is about 0 dBmV. Cable TV uses 75 Ω coaxial cable, so 0 dBmV corresponds to -78.75 dBW (-48.75 dBm) or ~13 nW.

dB μ V or dBuV

dB(1 μ V_{RMS}) — voltage relative to 1 microvolt. Widely used in television and aerial amplifier specifications. 60 dB μ V = 0 dBmV.

3. Properties of Symmetrical Networks and Characteristic impedance of Symmetrical Networks

A two-port network (a kind of four-terminal network or quadripole) is an electrical circuit or device with two pairs of terminals connected together internally by an electrical network. Two terminals constitute a port if they satisfy the essential requirement known as the port condition: the same current must enter and leave a port. Examples include small-signal models for transistors (such as the hybrid- π model), filters and matching networks. The analysis of passive two-port networks is an outgrowth of reciprocity theorems first derived by Lorentz^[3].

A two-port network makes possible the isolation of either a complete circuit or part of it and replacing it by its characteristic parameters. Once this is done, the isolated part of the circuit becomes a "black box" with a set of distinctive properties, enabling us to abstract away its specific physical buildup, thus simplifying analysis. Any linear circuit with four terminals can be transformed into a two-port network provided that it does not contain an independent source and satisfies the port conditions.

There are a number of alternative sets of parameters that can be used to describe a linear two-port network, the usual sets are respectively called z , y , h , g , and ABCD parameters, each described individually below. These are all limited to linear networks since an underlying assumption of their derivation is that any given circuit condition is a linear superposition of various short-circuit and open circuit conditions. They are usually expressed in matrix notation, and they establish relations between the variables

$V_1 \stackrel{\text{def}}{=} \text{Input voltage}$

$V_2 \stackrel{\text{def}}{=} \text{Output voltage}$

$I_1 \stackrel{\text{def}}{=} \text{Input current}$

$I_2 \stackrel{\text{def}}{=} \text{Output current}$

These current and voltage variables are most useful at low-to-moderate frequencies. At high frequencies (e.g., microwave frequencies), the use of power and energy variables is more appropriate, and the two-port current–voltage

an approach based upon scattering parameters.

The terms four-terminal network and quadripole (not to be confused with quadrupole) are also used, the latter particularly in more mathematical treatments although the term is becoming archaic. However, a pair of terminals can be called a port only if the current entering one terminal is equal to the current leaving the other; this definition is called the port condition. A four-terminal network can only be properly called a two-port when the terminals are connected to the external circuitry in two pairs both meeting the port condition.

4. voltage and current ratios

In order to simplify calculations, sinusoidal voltage and current waves are commonly represented as complex-valued functions of time denoted as \tilde{V} and \tilde{I} .^{[7][8]}

$$\tilde{V} = V_0 e^{j(\omega t + \phi_V)}$$

$$\tilde{I} = I_0 e^{j(\omega t + \phi_I)}$$

Impedance is defined as the ratio of these quantities.

$$\tilde{Z} = \frac{\tilde{V}}{\tilde{I}}$$

Substituting these into Ohm's law we have

$$\begin{aligned} V_0 e^{j(\omega t + \phi_V)} &= I_0 e^{j(\omega t + \phi_I)} Z e^{j\theta} \\ &= I_0 Z e^{j(\omega t + \phi_I + \theta)} \end{aligned}$$

Noting that this must hold for all t , we may equate the magnitudes and phases to obtain

$$V_0 = I_0 Z$$

$$\phi_V = \phi_I + \theta$$

The magnitude equation is the familiar Ohm's law applied to the voltage and current amplitudes, while the second equation defines the phase relationship.

Validity of complex representation

This representation using complex exponentials may be justified by noting that (by

Euler's formula):

$$\cos(\omega t + \phi) = \frac{1}{2} [e^{j(\omega t + \phi)} + e^{-j(\omega t + \phi)}]$$

i.e. a real-valued sinusoidal function (which may represent our voltage or current waveform) may be broken into two complex-valued functions. By the principle of superposition, we may analyse the behaviour of the sinusoid on the left-hand side by analysing the behaviour of the two complex terms on the right-hand side. Given the symmetry, we only need to perform the analysis for one right-hand term; the results will be identical for the other. At the end of any calculation, we may return to real-valued sinusoids by further noting that

$$\cos(\omega t + \phi) = \Re\{e^{j(\omega t + \phi)}\}$$

In other words, we simply take the real part of the result.

Phasors

A phasor is a constant complex number, usually expressed in exponential form, representing the complex amplitude (magnitude and phase) of a sinusoidal function of time. Phasors are used by electrical engineers to simplify computations involving sinusoids, where they can often reduce a differential equation problem to an algebraic one.

The impedance of a circuit element can be defined as the ratio of the phasor voltage across the element to the phasor current through the element, as determined by the relative amplitudes and phases of the voltage and current. This is identical to the definition from Ohm's law given above, recognising that the factors of $e^{j\omega t}$ cancel

5. Propagation constant

The propagation constant of an electromagnetic wave is a measure of the change undergone by the amplitude of the wave as it propagates in a given direction. The

quantity being measured can be the voltage or current in a circuit or a field vector such as electric field strength or flux density. The propagation constant itself measures change per metre but is otherwise dimensionless.

The propagation constant is expressed logarithmically, almost universally to the base e , rather than the more usual base 10 used in telecommunications in other situations. The quantity measured, such as voltage, is expressed as a sinusoidal phasor. The phase of the sinusoid varies with distance which results in the propagation constant being a complex number, the imaginary part being caused by the phase change.

Alternative names

The term propagation constant is somewhat of a misnomer as it usually varies strongly with ω . It is probably the most widely used term but there are a large variety of alternative names used by various authors for this quantity. These include, transmission parameter, transmission function, propagation parameter, propagation coefficient and transmission constant. In plural, it is usually implied that α and β are being referenced separately but collectively as in transmission parameters, propagation parameters, propagation coefficients, transmission constants and secondary coefficients. This last occurs in transmission line theory, the term secondary being used to contrast to the primary line coefficients. The primary coefficients being the physical properties of the line; R,C,L and G, from which the secondary coefficients may be derived using the telegrapher's equation. Note that, at least in the field of transmission lines, the term transmission coefficient has a different meaning despite the similarity of name. Here it is the

corollary of reflection coefficient.

Definition

The propagation constant, symbol γ , for a given system is defined by the ratio of the amplitude at the source of the wave to the amplitude at some distance x , such that,

$$\frac{A_0}{A_x} = e^{\gamma x}$$

Since the propagation constant is a complex quantity we can write;

$$\gamma = \alpha + i\beta$$

where

α , the real part, is called the attenuation constant

β , the imaginary part, is called the phase constant

That β does indeed represent phase can be seen from Euler's formula;

$$e^{i\theta} = \cos\theta + i\sin\theta$$

which is a sinusoid which varies in phase as θ varies but does not vary in amplitude because;

$$|e^{i\theta}| = \sqrt{\cos^2\theta + \sin^2\theta} = 1$$

The reason for the use of base e is also now made clear. The imaginary phase constant, $i\beta$, can be added directly to the attenuation constant, α , to form a single complex number that can be handled in one mathematical operation provided they are to the same base. Angles measured in radians require base e, so the attenuation is likewise in base e.

For a copper transmission line, the propagation constant can be calculated from the primary line coefficients by means of the relationship;

$$\gamma = \sqrt{ZY}$$

where;

$Z = R + i\omega L$, the series impedance of the line per metre and,

$Y = G + i\omega C$, the shunt admittance of the line per metre.

Attenuation constant

In telecommunications, the term attenuation constant, also called attenuation parameter or coefficient, is the attenuation of an electromagnetic wave propagating through a medium per unit distance from the source. It is the real part of the propagation constant and is measured in nepers per metre. A neper is approximately 8.7dB. Attenuation constant can be defined by the amplitude ratio;

$$\left| \frac{A_0}{A_x} \right| = e^{\alpha x}$$

The propagation constant per unit length is defined as the natural logarithmic of ratio of the sending end current or voltage to the receiving end current or voltage.

Copper lines

The attenuation constant for copper (or any other conductor) lines can be calculated from the primary line coefficients as shown above. For a line meeting the distortionless condition, with a conductance G in the insulator, the attenuation constant is given by;

$$\alpha = \sqrt{RG}$$

however, a real line is unlikely to meet this condition without the addition of loading coils and, furthermore, there are some decidedly non-linear effects operating on the primary "constants" which cause a frequency dependence of the loss. There are two main components to these losses, the metal loss and the dielectric loss.

The loss of most transmission lines are dominated by the metal loss, which causes a frequency dependency due to finite conductivity of metals, and the skin effect inside a conductor. The skin effect causes R along the conductor to be approximately dependent on frequency according to;

$$R \propto \sqrt{\omega}$$

Losses in the dielectric depend on the loss tangent ($\tan\delta$) of the material, which depends inversely on the wavelength of the signal and is directly proportional to the frequency.

$$\alpha_d = \frac{\pi \sqrt{\epsilon_r}}{\lambda} \tan \delta$$

Optical fibre

The attenuation constant for a particular propagation mode in an optical fiber, the real part of the axial propagation constant.

Phase constant

In electromagnetic theory, the phase constant, also called phase change constant, parameter or coefficient is the imaginary component of the propagation constant for a plane wave. It represents the change in phase per metre along the path travelled by the wave at any instant and is equal to the angular wavenumber of the wave. It is represented by the symbol β and is measured in units of radians per metre.

From the definition of angular wavenumber;

$$k = \frac{2\pi}{\lambda} = \beta$$

This quantity is often (strictly speaking incorrectly) abbreviated to wavenumber. Properly, wavenumber is given by,

$$\tilde{\nu} = 1/\lambda$$

which differs from angular wavenumber only by a constant multiple of 2π , in the same way that angular frequency differs from frequency.

For a transmission line, the Heaviside condition of the telegrapher's equation tells us that the wavenumber must be proportional to frequency for the transmission of the wave to be undistorted in the time domain. This includes, but is not limited to, the ideal case of a lossless line. The reason for this condition can be seen by considering that a useful signal is composed of many different wavelengths in the frequency domain. For there to be no distortion of the waveform, all these waves must travel at the same velocity so that they arrive at the far end of the line at the same time as a group. Since wave phase velocity is given by;

$$v_p = \frac{\lambda}{T} = \frac{f}{\tilde{\nu}} = \frac{\omega}{\beta}$$

it is proved that β is required to be proportional to ω . In terms of primary coefficients of the line, this yields from the telegrapher's equation for a distortionless line the condition;

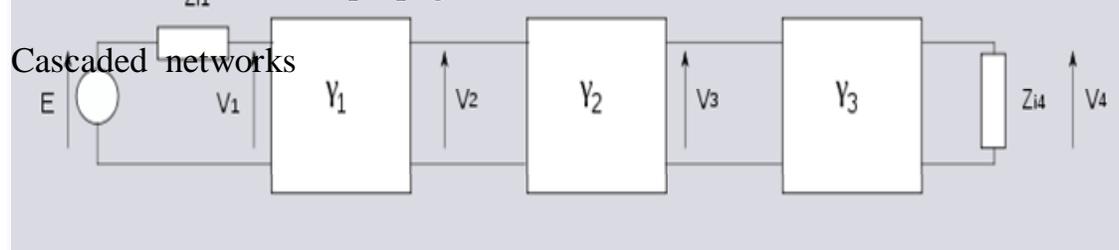
$$\beta = \omega\sqrt{LC}$$

However, practical lines can only be expected to approximately meet this condition over a limited frequency band.

6. Filters

The term propagation constant or propagation function is applied to filters and other two-port networks used for signal processing. In these cases, however, the attenuation and phase coefficients are expressed in terms of nepers and radians per network section rather than per metre. Some authors make a distinction between per metre measures (for which "constant" is used) and per section measures (for which "function" is used).

The propagation constant is a useful concept in filter design which invariably uses a cascaded section topology. In a cascaded topology, the propagation constant, attenuation constant and phase constant of individual sections may be simply added to find the total propagation constant etc.



Three networks with arbitrary propagation constants and impedances connected in

cascade. The Z_i terms represent image impedance and it is assumed that connections are between matching image impedances.

The ratio of output to input voltage for each network is given by,

$$\frac{V_1}{V_2} = \sqrt{\frac{Z_{I1}}{Z_{I2}}} e^{\gamma_1}$$

$$\frac{V_2}{V_3} = \sqrt{\frac{Z_{I2}}{Z_{I3}}} e^{\gamma_2}$$

$$\frac{V_3}{V_4} = \sqrt{\frac{Z_{I3}}{Z_{I4}}} e^{\gamma_3}$$

The terms $\sqrt{\frac{Z_{In}}{Z_{Im}}}$ are impedance scaling terms^[3] and their use is explained in the image impedance article.

The overall voltage ratio is given by,

$$\frac{V_1}{V_4} = \frac{V_1}{V_2} \cdot \frac{V_2}{V_3} \cdot \frac{V_3}{V_4} = \sqrt{\frac{Z_{I1}}{Z_{I4}}} e^{\gamma_1 + \gamma_2 + \gamma_3}$$

Thus for n cascaded sections all having matching impedances facing each other, the overall propagation constant is given by,

$$\gamma_{Tot} = \gamma_1 + \gamma_2 + \gamma_3 + \dots + \gamma_n$$

7. Filter fundamentals – Pass and Stop bands.

filters of all types are required in a variety of applications from audio to RF and across the whole spectrum of frequencies. As such RF filters form an important element within a variety of scenarios, enabling the required frequencies to be passed through the circuit, while rejecting those that are not needed.

The ideal filter, whether it is a low pass, high pass, or band pass filter will exhibit no loss within the pass band, i.e. the frequencies below the cut off frequency. Then above this frequency in what is termed the stop band the filter will reject all signals.

In reality it is not possible to achieve the perfect pass filter and there is always some loss within the pass band, and it is not possible to achieve infinite rejection in the stop band. Also there is a transition between the pass band and the stop band,

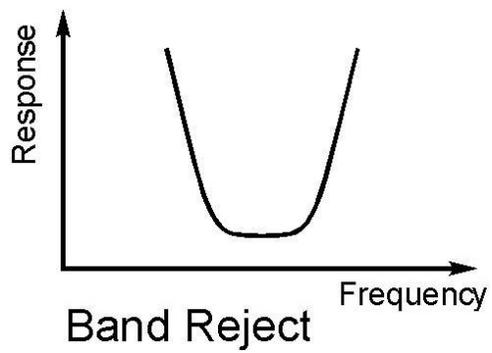
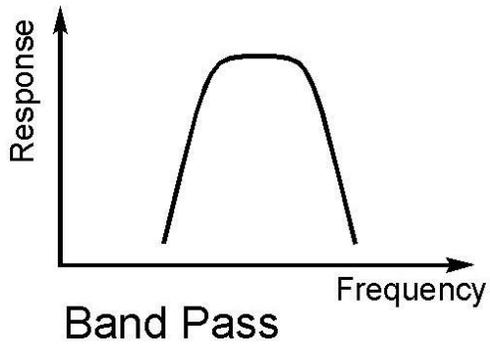
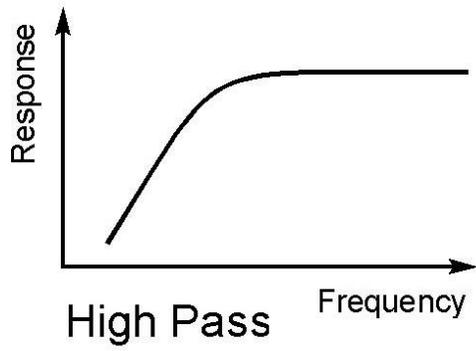
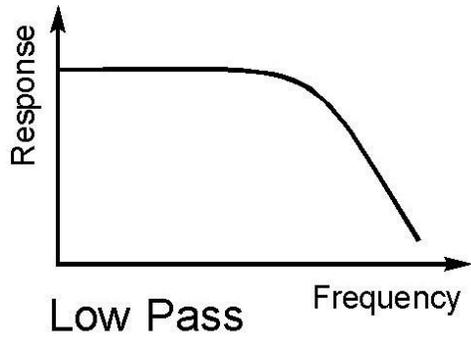
where the response curve falls away, with the level of rejection rises as the frequency moves from the pass band to the stop band.

Basic types of RF filter

There are four types of filter that can be defined. Each different type rejects or accepts signals in a different way, and by using the correct type of RF filter it is possible to accept the required signals and reject those that are not wanted. The four basic types of RF filter are:

- Low pass filter
- High pass filter
- Band pass filter
- Band reject filter

As the names of these types of RF filter indicate, a low pass filter only allows frequencies below what is termed the cut off frequency through. This can also be thought of as a high reject filter as it rejects high frequencies. Similarly a high pass filter only allows signals through above the cut off frequency and rejects those below the cut off frequency. A band pass filter allows frequencies through within a given pass band. Finally the band reject filter rejects signals within a certain band. It can be particularly useful for rejecting a particular unwanted signal or set of signals falling within a given bandwidth.



filter frequencies

A filter allows signals through in what is termed the pass band. This is the band of frequencies below the cut off frequency for the filter.

The cut off frequency of the filter is defined as the point at which the output level from the filter falls to 50% (-3 dB) of the in band level, assuming a constant input level. The cut off frequency is sometimes referred to as the half power or -3 dB frequency.

The stop band of the filter is essentially the band of frequencies that is rejected by the filter. It is taken as starting at the point where the filter reaches its required level of rejection.

Filter classifications

Filters can be designed to meet a variety of requirements. Although using the same basic circuit configurations, the circuit values differ when the circuit is designed to meet different criteria. In band ripple, fastest transition to the ultimate roll off,

highest out of band rejection are some of the criteria that result in different circuit values. These different filters are given names, each one being optimised for a

different element of performance. Three common types of filter are given below:

- Butterworth: This type of filter provides the maximum in band flatness.
- Bessel: This filter provides the optimum in-band phase response and therefore also provides the best step response.
- Chebychev: This filter provides fast roll off after the cut off frequency is reached. However this is at the expense of in band ripple. The more in band ripple that can be tolerated, the faster the roll off.
- Elliptical: This has significant levels of in band and out of band ripple, and as expected the higher the degree of ripple that can be tolerated, the steeper it reaches its ultimate roll off.

Summary

RF filters are widely used in RF design and in all manner of RF and analogue circuits in general. As they allow though only particular frequencies or bands of frequencies, they are an essential tool for the RF design engineer.

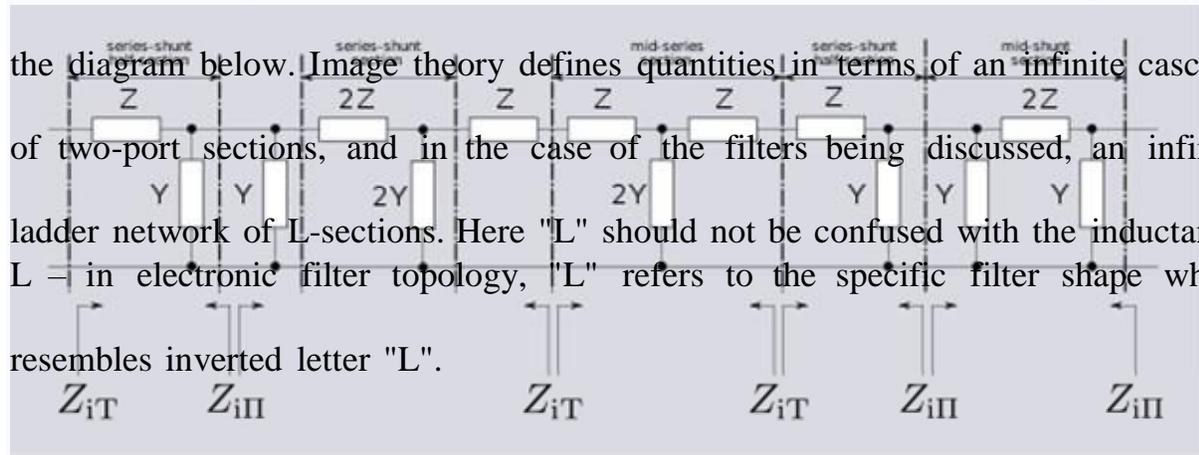
8. Constant k filter

Constant k filters, also k-type filters, are a type of electronic filter designed using the image method. They are the original and simplest filters produced by this methodology and consist of a ladder network of identical sections of passive components. Historically, they are the first filters that could approach the ideal filter frequency response to within any prescribed limit with the addition of a sufficient number of sections. However, they are rarely considered for a modern design, the principles behind them having been superseded by other methodologies which are more accurate in their prediction of filter response.

Terminology

Some of the impedance terms and section terms used in this article are pictured in the diagram below.

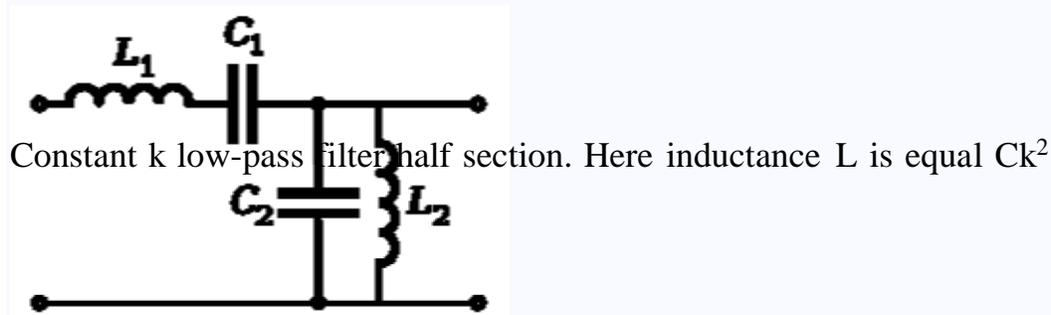
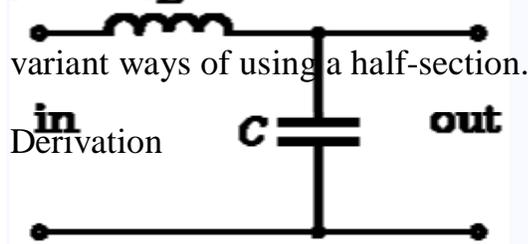
Image theory defines quantities in terms of an infinite cascade of two-port sections, and in the case of the filters being discussed, an infinite ladder network of L-sections. Here "L" should not be confused with the inductance $L = \text{---}$ in electronic filter topology, "L" refers to the specific filter shape which resembles inverted letter "L".



The sections of the hypothetical infinite filter are made of series elements having

impedance $2Z$ and shunt elements with admittance $2Y$. The factor of two is introduced for mathematical convenience, since it is usual to work in terms of half-

sections where it disappears. The image impedance of the input and output port of a section will generally not be the same. However, for a mid-series section (that is, a section from halfway through a series element to halfway through the next series element) will have the same image impedance on both ports due to symmetry. This image impedance is designated Z_{iT} due to the "T" topology of a mid-series section. Likewise, the image impedance of a mid-shunt section is designated $Z_{i\Pi}$ due to the "II" topology. Half of such a "T" or "II" section is called a half-section, which is also an L-section but with half the element values of the full L-section. The image impedance of the half-section is dissimilar on the input and output ports: on the side presenting the series element it is equal to the mid-series Z_{iT} , but on the side presenting the shunt element it is equal to the mid-shunt $Z_{i\Pi}$. There are thus two



Constant k band-pass filter half section.
 $L_1 = C_2k^2$ and $L_2 = C_1k^2$

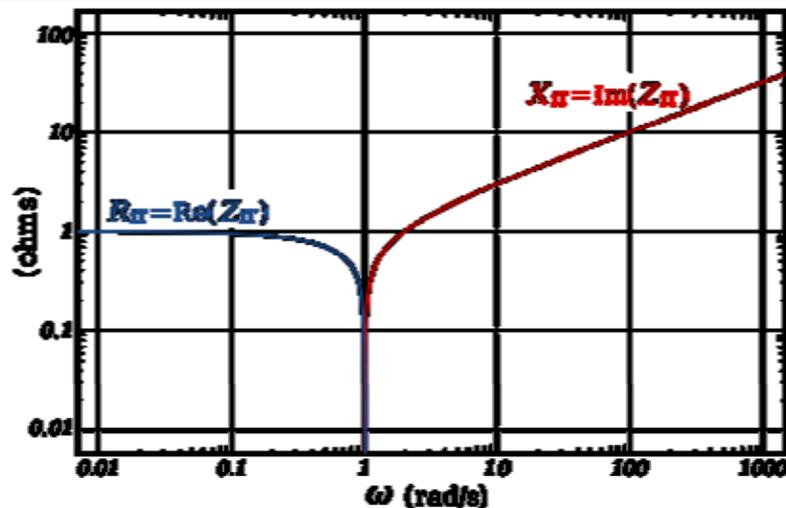


Image impedance Z_{iT} of a constant k prototype low-pass filter is plotted vs. frequency ω . The impedance is purely resistive (real) below ω_c , and purely reactive (imaginary) above ω_c .

The building block of constant k filters is the half-section "L" network, composed of a series impedance Z , and a shunt admittance Y . The "k" in "constant k" is the value given by,^[6]

$$k^2 = \frac{Z}{Y}$$

Thus, k will have units of impedance, that is, ohms. It is readily apparent that in order for k to be constant, Y must be the dual impedance of Z . A physical interpretation of k can be given by observing that k is the limiting value of Z_i as the size of the section (in terms of values of its components, such as inductances, capacitances, etc.) approaches zero, while keeping k at its initial value. Thus, k is the characteristic impedance, Z_0 , of the transmission line that would be formed by these infinitesimally small sections. It is also the image impedance of the section at resonance, in the case of band-pass filters, or at $\omega = 0$ in the case of low-pass filters.^[7] For example, the pictured low-pass half-section has

$$k = \sqrt{\frac{i\omega L}{i\omega C}} = \sqrt{\frac{L}{C}}$$

lements L and C can be made arbitrarily small while retaining the same value of k . Z and Y however, are both approaching zero, and from the formulae (below) for image impedances,

$$\lim_{Z,Y \rightarrow 0} Z_i = k$$

Image impedance

The image impedances of the section are given by^[8]

$$Z_{iT}^2 = Z^2 + k^2$$

and

$$\frac{1}{Z_{iII}^2} = Y_{iII}^2 = Y^2 + \frac{1}{k^2}$$

Provided that the filter does not contain any resistive elements, the image impedance in the pass band of the filter is purely real and in the stop band it is purely imaginary. For example, for the pictured low-pass half-section,^[9]

$$Z_{iT}^2 = -(\omega L)^2 + \frac{L}{C}$$

The transition occurs at a cut-off frequency given by

$$\omega_c = \frac{1}{\sqrt{LC}}$$

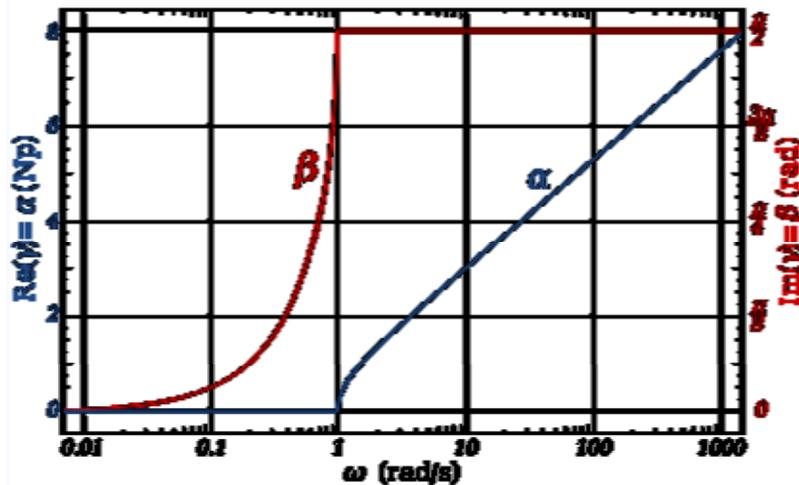
Below this frequency, the image impedance is real,

$$Z_{iT} = L\sqrt{\omega_c^2 - \omega^2}$$

Above the cut-off frequency the image impedance is imaginary,

$$Z_{iT} = iL\sqrt{\omega^2 - \omega_c^2}$$

Transmission parameters



The transfer function of a constant k prototype low-pass filter for a single half-section showing attenuation in nepers and phase change in radians.

See also: Image impedance#Transfer function

The transmission parameters for a general constant k half-section are given by^[10]

$$\gamma = \sinh^{-1} \frac{Z}{k}$$

and for a chain of n half-sections

$$\gamma_n = n\gamma$$

For the low-pass L-shape section, below the cut-off frequency, the transmission parameters are given by^[8]

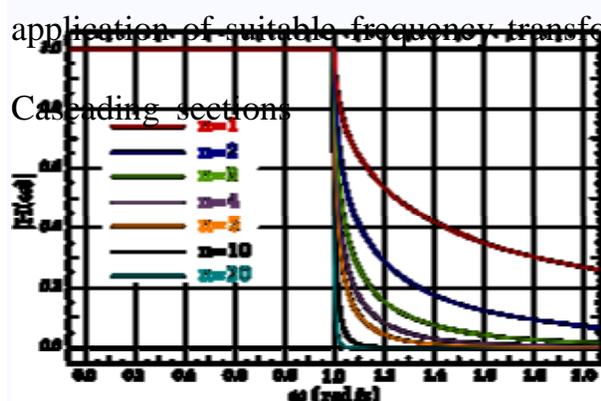
$$\gamma = \alpha + i\beta = 0 + i \sin^{-1} \frac{\omega}{\omega_c}$$

That is, the transmission is lossless in the pass-band with only the phase of the signal changing. Above the cut-off frequency, the transmission parameters are:^[8]

$$\gamma = \alpha + i\beta = \cosh^{-1} \frac{\omega}{\omega_c} + i \frac{\pi}{2}$$

Prototype transformations

The presented plots of image impedance, attenuation and phase change correspond to a low-pass prototype filter section. The prototype has a cut-off frequency of $\omega_c = 1$ rad/s and a nominal impedance $k = 1 \Omega$. This is produced by a filter half-section with inductance $L = 1$ henry and capacitance $C = 1$ farad. This prototype can be impedance scaled and frequency scaled to the desired values. The low-pass prototype can also be transformed into high-pass, band-pass or band-stop types by application of suitable frequency transformations.^[11]

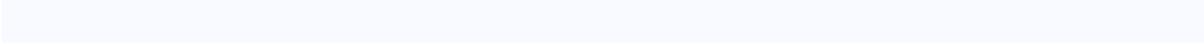


Gain response, $H(\omega)$ for a chain of n low-pass constant- k filter half-sections.

Several L-shape half-sections may be cascaded to form a composite filter. Like impedance must always face like in these combinations. There are therefore two circuits that can be formed with two identical L-shaped half-sections. Where a port of image impedance Z_{iT} faces another Z_{iT} , the section is called a Π section. Where $Z_{i\Pi}$ faces $Z_{i\Pi}$ the section so formed is a T section. Further additions of half-sections to either of these section forms a ladder network which may start and end with series or shunt elements.^[12]

It should be borne in mind that the characteristics of the filter predicted by the image method are only accurate if the section is terminated with its image impedance. This is usually not true of the sections at either end, which are usually

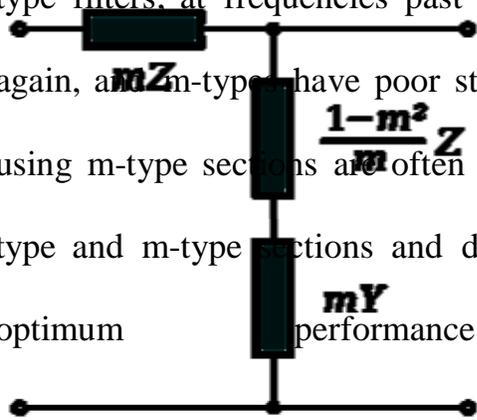
terminated with a fixed resistance. The further the section is from the end of the filter, the more accurate the prediction will become, since the effects of the terminating impedances are masked by the intervening sections.^[13]



9. m-derived filter

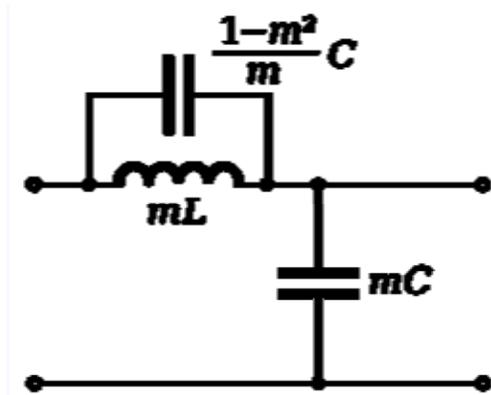
m-derived filters or m-type filters are a type of electronic filter designed using the image method. They were invented by Otto Zobel in the early 1920s.^[1] This filter type was originally intended for use with telephone multiplexing and was an improvement on the existing constant k type filter.^[2] The main problem being addressed was the need to achieve a better match of the filter into the terminating impedances. In general, all filters designed by the image method fail to give an exact match, but the m-type filter is a big improvement with suitable choice of the parameter m. The m-type filter section has a further advantage in that there is a rapid transition from the cut-off frequency of the pass band to a pole of attenuation just inside the stop band. Despite these advantages, there is a drawback with m-

type filters; at frequencies past the pole of attenuation, the response starts to rise again, and m-type have poor stop band rejection. For this reason, filters designed using m-type sections are often designed as composite filters with a mixture of k-type and m-type sections and different values of m at different points to get the optimum performance from both types.^[3]



Derivation

m-derived series general filter half section.



m-derived shunt low-pass filter half section.

$$C = \frac{L}{R_0^2}$$

The building block of m-derived filters, as with all image impedance filters, is the "L" network, called a half-section and composed of a series impedance Z, and a shunt admittance Y. The m-derived filter is a derivative of the constant k filter. The starting point of the design is the values of Z and Y derived from the constant k prototype and are given by

$$k^2 = \frac{Z}{Y}$$

where k is the nominal impedance of the filter, or R_0 . The designer now multiplies Z and Y by an arbitrary constant m ($0 < m < 1$). There are two different kinds of m-derived section; series and shunt. To obtain the m-derived series half section, the designer determines the impedance that must be added to $1/mY$ to make the image impedance Z_{iT} the same as the image impedance of the original constant k section. From the general formula for image impedance, the additional impedance required can be shown to be^[9]

$$\frac{1 - m^2}{m} Z.$$

To obtain the m-derived shunt half section, an admittance is added to $1/mZ$ to make the image impedance $Z_{i\Pi}$ the same as the image impedance of the original half section. The additional admittance required can be shown to be^[10]

$$\frac{1 - m^2}{m} Y.$$

The general arrangements of these circuits are shown in the diagrams to the right along with a specific example of a low pass section.

A consequence of this design is that the m-derived half section will match a k-type section on one side only. Also, an m-type section of one value of m will not match another m-type section of another value of m except on the sides which offer the Z_i of the k-type.^[11]

Operating frequency

For the low-pass half section shown, the cut-off frequency of the m-type is the same as the k-type and is given by

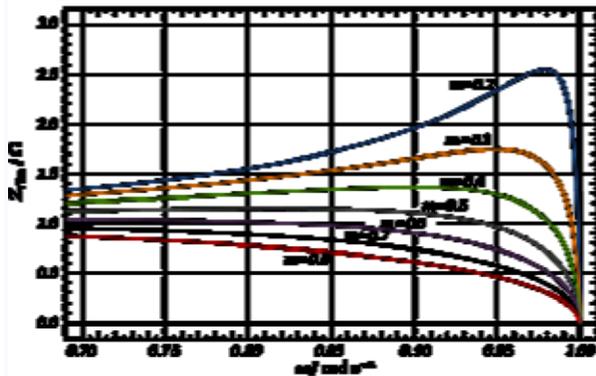
$$\omega_c = \frac{1}{\sqrt{LC}}.$$

The pole of attenuation occurs at;

$$\omega_\infty = \frac{\omega_c}{\sqrt{1 - m^2}}.$$

From this it is clear that smaller values of m will produce ω_∞ closer to the cut-off frequency and hence will have a sharper cut-off. Despite this cut-off, it also brings the unwanted stop band response of the m-type closer to the cut-off frequency, making it more difficult for this to be filtered with subsequent sections. The value of m chosen is usually a compromise between these conflicting requirements. There is also a practical limit to how small m can be made due to the inherent resistance of the inductors. This has the effect of causing the pole of attenuation to be less deep (that is, it is no longer a genuinely infinite pole) and the slope of cut-off to be less steep. This effect becomes more marked as ω_∞ is brought closer to ω_c , and there ceases to be

ω_c
Image impedance



m-derived prototype shunt low-pass filter Z_{iTm} image impedance for various values of m. Values below cut-off frequency only shown for clarity.

The following expressions for image impedances are all referenced to the low-pass prototype section. They are scaled to the nominal impedance $R_0 = 1$, and the frequencies in those expressions are all scaled to the cut-off frequency $\omega_c = 1$.

Series sections

The image impedances of the series section are given by^[14]

$$Z_{iT} = \sqrt{1 - \omega^2}$$

and is the same as that of the constant k section

$$Z_{i\Pi m} = \frac{1 - (\omega/\omega_\infty)^2}{\sqrt{1 - \omega^2}}$$

Shunt sections

The image impedances of the shunt section are given by^[11]

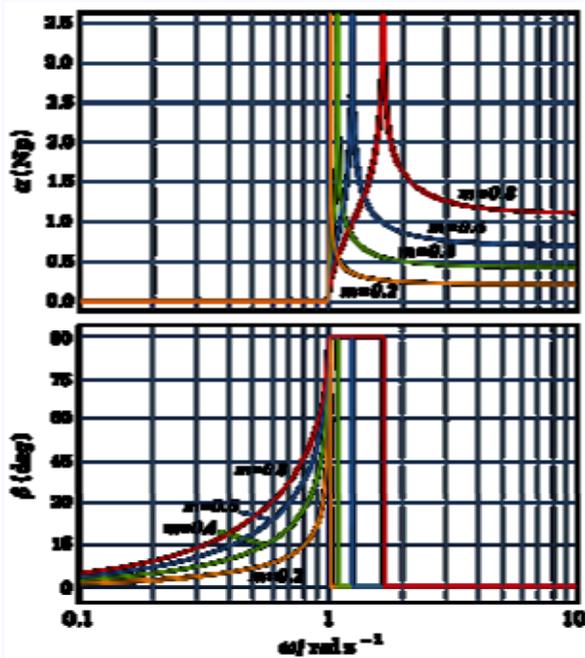
$$Z_{i\Pi} = \frac{1}{\sqrt{1 - \omega^2}}$$

and is the same as that of the constant k section

$$Z_{iTm} = \frac{\sqrt{1 - \omega^2}}{1 - (\omega/\omega_\infty)^2}$$

As with the k-type section, the image impedance of the m-type low-pass section is purely real below the cut-off frequency and purely imaginary above it. From the chart it can be seen that in the passband the closest impedance match to a constant pure resistance termination occurs at approximately $m = 0.6$.^[14]

Transmission parameters



m-Derived low-pass filter transfer function for a single half-section

For an m-derived section in general the transmission parameters for a half-section are given by^[14]

$$\gamma = \sinh^{-1} \frac{mZ}{\sqrt{k^2 + (1 - m^2)Z^2}}$$

and for n half-sections

$$\gamma_n = n\gamma$$

For the particular example of the low-pass L section, the transmission parameters solve differently in three frequency bands.^[14]

For $0 < \omega < \omega_c$ the transmission is lossless:

$$\gamma = \alpha + i\beta = 0 + i\frac{1}{2} \cos^{-1} \left(1 - \frac{2m^2}{\left(\frac{\omega_c}{\omega}\right)^2 - \left(\frac{\omega_c}{\omega_\infty}\right)^2} \right)$$

For $\omega_c < \omega < \omega_\infty$ the transmission parameters are

$$\gamma = \alpha + i\beta = \frac{1}{2} \cosh^{-1} \left(\frac{2m^2}{\left(\frac{\omega_c}{\omega}\right)^2 - \left(\frac{\omega_c}{\omega_\infty}\right)^2} - 1 \right) + i\frac{\pi}{2}$$

For $\omega_\infty < \omega < \infty$ the transmission parameters are

$$\gamma = \alpha + i\beta = \frac{1}{2} \cosh^{-1} \left(1 - \frac{2m^2}{\left(\frac{\omega_c}{\omega}\right)^2 - \left(\frac{\omega_c}{\omega_\infty}\right)^2} \right) + i0$$

Prototype transformations

The plots shown of image impedance, attenuation and phase change are the plots of a low-pass prototype filter section. The prototype has a cut-off frequency of $\omega_c = 1$ rad/s and a nominal impedance $R_0 = 1 \Omega$. This is produced by a filter half-section where $L = 1$ henry and $C = 1$ farad. This prototype can be impedance scaled and frequency scaled to the desired values. The low-pass prototype can also be transformed into high-pass, band-pass or band-stop types by application of suitable frequency transformations.^[15]

Cascading sections

Several L half-sections may be cascaded to form a composite filter. Like impedance must always face like in these combinations. There are therefore two circuits that can be formed with two identical L half-sections. Where Z_{iT} faces Z_{iT} , the section is called a Π section. Where $Z_{i\Pi}$ faces $Z_{i\Pi}$ the section formed is a T section. Further additions of half-sections to either of these forms a ladder network which may start and end with series or shunt elements.^[16]

It should be born in mind that the characteristics of the filter predicted by the image method are only accurate if the section is terminated with its image impedance. This is usually not true of the sections at either end which are usually terminated with a fixed resistance. The further the section is from the end of the filter, the more accurate the prediction will become since the effects of the terminating impedances are masked by the intervening sections. It is usual to provide half half-sections at the ends of the filter with $m = 0.6$ as this value gives the flattest Z_i in the passband and hence the best match in to a resistive termination.^[17]

10. Crystal filter

A crystal filter is a special form of quartz crystal used in electronics systems, in particular communications devices. It provides a very precisely defined centre frequency and very steep bandpass characteristics, that is a very high Q factor—far higher than can be obtained with conventional lumped circuits.

A crystal filter is very often found in the intermediate frequency (IF) stages of high-quality radio receivers. Cheaper sets may use ceramic filters (which also exploit the piezoelectric effect), or tuned LC circuits. The use of a fixed IF stage frequency allows a crystal filter to be used because it has a very precise fixed frequency.

The most common use of crystal filters, is at frequencies of 9 MHz or 10.7 MHz to provide selectivity in communications receivers, or at higher frequencies as a roofing filter in receivers using up-conversion.

Ceramic filters tend to be used at 10.7 MHz to provide selectivity in broadcast FM receivers, or at a lower frequency (455 kHz) as the second intermediate frequency filters in a communication receiver. Ceramic filters at 455 kHz can achieve similar bandwidths to crystal filters at 10.7 MHz.

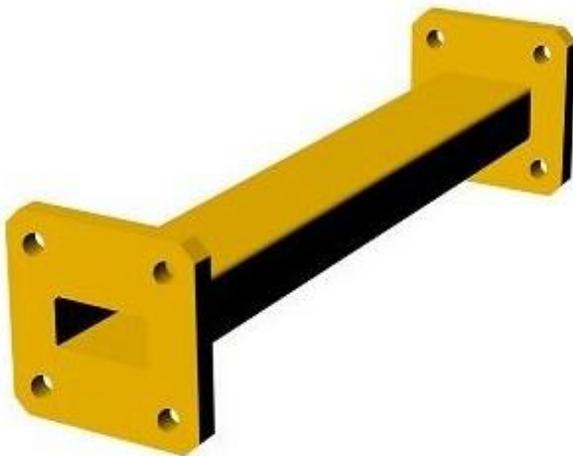
UNIT V

WAVEGUIDES

Waveguides are basically a device ("a guide") for transporting electromagnetic energy from one region to another. Typically, waveguides are hollow metal tubes (often rectangular or circular in cross section). They are capable of directing power precisely to where it is needed, can handle large amounts of power and function as a high-pass filter.

The waveguide acts as a high pass filter in that most of the energy above a certain frequency (the cutoff frequency) will pass through the waveguide, whereas most of the energy that is below the cutoff frequency will be attenuated by the waveguide. Waveguides are often used at microwave frequencies (greater than 300 MHz, with 8 GHz and above being more common).

Waveguides are wideband devices, and can carry (or transmit) either power or communication signals. An example of a hollow metal rectangular waveguide is shown in the following figure.



Waveguides can bend if the desired application requires it, as shown in the following Figure.



The above waveguides can be used with waveguide to coaxial cable adapters, as shown in the next Figure:



We now know what a waveguide is. Lets examine metal cavities with a rectangular cross section, as shown in Figure 1. Assume the waveguide is filled with vaccuum, air or some dielectric with the permeability given by μ and the permittivity given by ϵ .

The waveguide has a width a in the x -direction, and a height b in the y -direction, with $a > b$. The z -axis is the direction in which the waveguide is to carry power.

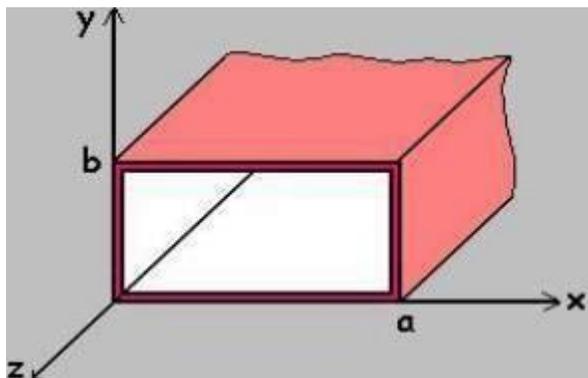


Figure 1. Cross section of a waveguide with long dimension a and short dimension b .

On this page, I'm going to give the general "rules" for waveguides. That is, I'll give the equations for key parameters and let you know what the parameters mean. On the next page, we'll go into the mathematical derivation (which you would do in engineering graduate school), but you can get away with not knowing all that math if you don't want to know it.

First and possibly most importantly, this waveguide has a cutoff frequency, f_c . The cutoff frequency is the frequency at which all lower frequencies are attenuated by the waveguide, and above the cutoff frequency all higher frequencies propagate

within the waveguide. The cutoff frequency defines the high-pass filter characteristic of the waveguide: above this frequency, the waveguide passes power, below this frequency the waveguide attenuates or blocks power.

The cutoff frequency depends on the shape and size of the cross section of the waveguide. The larger the waveguide is, the lower the cutoff frequency for that waveguide is. The formula for the cutoff frequency of a rectangular cross sectioned waveguide is given by:

$$f_c = \frac{1}{2a\sqrt{\mu\epsilon}} = \frac{c}{2a}$$

[Rectangular Cross Section]

www.antenna-theory.com

In the above, c is the speed of light within the waveguide, μ is the permeability of the material that fills the waveguide, and ϵ is the permittivity of the material that fills the waveguide. Note that the cutoff frequency is independent of the short length b of the waveguide.

The cutoff frequency for a waveguide with a circular cross section of radius a is given by:

Due to Maxwell's Equations, the fields within the waveguide always have a specific "form" or "waveshape" to them - these are called modes. Assume the waveguide is oriented such that the energy is to be transmitted along the waveguide axis, the z -axis. The modes are classified as either TE ('transverse electric' - which indicates that the E-field is orthogonal to the axis of the waveguide, so that $E_z=0$) or TM ('transverse magnetic' - which indicates that the H-field is orthogonal to the axis of the waveguide, so $H_z = 0$). The modes are further classified as TE_{ij} , where the i and j indicate the number of wave oscillations for a particular field direction in the long direction (dimension a in Figure 1) and short direction (dimension b in Figure 1), respectively.

Metal waveguides cannot support the TEM ('transverse electric and magnetic' - when E_z and H_z are zero) mode. There exists no solution to Maxwell's equations that also satisfy the required boundary conditions for this mode to occur.

Maxwell's Equations are not easy to solve. Hence, every math trick someone can think of will be used in order to make the analysis tractable. We'll start with discussing the electric vector potential, F . In a source-free region (i.e., an area through which waves propagate that is away from sources), we know that:

$$\nabla \cdot \mathbf{D} = 0$$

In the above, \mathbf{D} is the Electric Flux Density. If a vector quantity is divergenceless (as in the above), then it can be expressed as the curl of another quantity. This

means that we can write the solution for \mathbf{D} and the corresponding electric field \mathbf{E} as:

$$\mathbf{D} = -\nabla \times \mathbf{F}$$

www.antenna-theory.com

$$\mathbf{E} = \frac{-1}{\epsilon} \nabla \times \mathbf{F}$$

In the above, epsilon is the permittivity of the medium through which the wave

propagates. We are purely in the world of mathematics now. The quantity F is not physical, and is of little practical value. It is simply an aid in performing our mathematical manipulations.

It turns out that waves (or electromagnetic energy) can not propagate in a waveguide when both H_z and E_z are equal to zero. Hence, what field configurations that are allowed will be classified as either TM (Transverse Magnetic, in which $H_z=0$) and TE (Transverse Electric, in which $E_z=0$). The reason that waves cannot be TEM (Transverse Electromagnetic, $H_z=E_z=0$) will be shown towards the end of this derivation.

To perform our analysis, we'll assume that $E_z=0$ (i.e., we are looking at a TE mode or field configuration). In this case, working through Maxwell's equations, it can be shown that the E- and H- fields can be determined from the following equations:

$E_x = \frac{-1}{\epsilon} \frac{\partial F_z}{\partial y}$ $E_y = \frac{1}{\epsilon} \frac{\partial F_z}{\partial x}$ $E_z = 0$	$H_x = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 F_z}{\partial x \partial z}$ $H_y = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 F_z}{\partial y \partial z}$ $H_z = -j \frac{1}{\omega \mu \epsilon} \left(\frac{\partial^2}{\partial z^2} + k^2 \right) F_z$
--	---

Therefore, if we can find F_z (the z-component of the vector F), then we can find the E- and H- fields. In the above equation, k is the wavenumber.

Working through the math of Maxwell's Equations, it can be shown that in a source-free region, the vector potential F must satisfy the vector wave equation:

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \mathbf{F} = 0 \quad [1]$$

To break this equation down, we will look only at the z-component of the above

equation (that is, Fz). We will also assume that we are looking at a single frequency, so that the time dependence is assumed to be of the form given by (we are now using phasors to analyze the equation):

$$e^{j\omega t} = e^{j2\pi ft}$$

Then the equation [1] can be simplified as follows:

$$\nabla^2 F_z + k^2 F_z = 0$$

$$\frac{\partial^2 F_z}{\partial x^2} + \frac{\partial^2 F_z}{\partial y^2} + \frac{\partial^2 F_z}{\partial z^2} + k^2 F_z = 0 \quad [2]$$

To solve this equation, we will use the technique of separation of variables. Here we assume that the function $F_z(x, y, z)$ can be written as the product of three functions, each of a single variable. That is, we assume that:

$$F_z(x, y, z) = X(x)Y(y)Z(z) \quad [3]$$

(You might ask, how do we know that the separation of variables assumption above is valid? We don't - we just assume its correct, and if it solves the differential equation when we are done doing the analysis then the assumption is valid). Now we plug in our assumption for F_z (equation [3]) into equation [2], and we end up with:

$$\frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} = -k^2 \quad [4]$$

In the above equation, the prime represents the derivative with respect to the variable in the equation (for instance, Z' represents the derivative of the Z -function with respect to z). We will break up the variable k^2 into components (again, just to make our math easier):

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \omega^2 \mu \epsilon \quad [5]$$

Using equation [5] to breakdown equation [4], we can write:

$$X'' + k_x^2 X = 0$$

$$Y'' + k_y^2 Y = 0$$

[6]

The reason that the equations in [6] are valid is because they are only functions of independent variables - hence, each equation must hold for [5] to be true everywhere in the waveguide. Solving the above equations using ordinary differential equations theory, we get:

$$\begin{aligned} X(x) &= c_1 \cos(k_x x) + c_2 \sin(k_x x) \\ Y(y) &= c_3 \cos(k_y y) + c_4 \sin(k_y y) \\ Z(z) &= c_5 e^{jk_z z} + c_6 e^{-jk_z z} \end{aligned} \quad [7]$$

The form of the solution in the above equation is different for Z(z). The reason is that both forms (that for X and Y, and that for Z), are both equally valid solutions for the differential equations in equation [6]. However, the complex exponential typically represents travelling waves, and the [real] sinusoids represent standing waves. Hence, we choose the forms given in [7] for the solutions. No math rules are violated here; again, we are just choosing forms that will make our analysis easier.

For now, we can set $c_5=0$, because we want to analyze waves propagating in the +z-direction. The analysis is identical for waves propagating in the -z-direction, so this is fairly arbitrary. The solution for F_z can be written as:

$$F_z = [c_1 \cos(k_x x) + c_2 \sin(k_x x)] \{c_3 \cos(k_y y) + c_4 \sin(k_y y)\} c_6 e^{-jk_z z}$$

[8]

If you remember anything about differential equations, you know there needs to be some boundary conditions applied in order to determine the constants. Recalling our physics, we know that the ~~tangential~~ $J = \sigma E$ Electric fields at any perfect conductor must be zero (why? because _____), so if the conductivity approaches infinity

(perfect conductor),
then if the tangential E-field is not zero then the induced current would be infinite).

The tangential fields must be zero, so E_x must be zero when $y=0$ and when $y=b$

(see Figure 1 above), no matter what the value for y and z are. In addition, E_y must be zero when $x=0$ and when $x=a$ (independent of x and z). We will calculate E_x :

$$E_x = \frac{-1}{\epsilon} \frac{\partial F_z}{\partial y}$$

www.antenna-theory.com

$$= \frac{-c_6 k_y}{\epsilon} [c_1 \cos(k_x x) + c_2 \sin(k_x x)] \{-c_3 \sin(k_y y) + c_4 \cos(k_y y)\} e^{-jk_z z}$$

E_x is given by the above equation. The boundary condition given by

$$E_x(x, y=0, z)=0 \quad [9]$$

implies that c_4 must be equal to zero. This is the only way that boundary condition given in [9] will be true for all x and z positions. If you don't believe this, try to

show that it is incorrect. You will quickly determine that c_4 must be zero for the boundary condition in [9] to be satisfied everywhere it is required.

Next, the second boundary condition,

$$E_x(x, y=b, z)=0 \quad [10]$$

implies something very unique. The only way for the condition in [10] to be true for all values of x and z whenever $y=b$, we must have:

$$c_3 \sin(k_y b) = 0$$

If this is to be true everywhere, c_3 could be zero. However, if c_3 is zero (and we have already determined that c_4 is zero), then all of the fields would end up being zero, because the function $Y(y)$ in [7] would be zero everywhere. Hence, c_3 cannot be zero if we are looking for a nonzero solution. Hence, the only alternative is if

the above equation implies that:

$$c_3 \sin(k_y b) = 0 \rightarrow k_y = \frac{n\pi}{b}, \quad n = 1, 2, \dots$$

This last equation is fundamental to understanding waveguides. It states that the

www.padeepz.net

only solutions for $Y(y)$ function must end up being sinusoids, that an integer number of multiples of a half-wavelength. These are the only type of functions that

satisfy the differential equation in [6] and the required boundary conditions. This is an extremely important concept.

If we invoke our other two boundary conditions:

$$E_y(x=0, y, z)=0$$

$$E_y(x=a, y, z)=0$$

Then (using identical reasoning to that above), we can determine that $c_2=0$ and that:

$$k_x = \frac{m\pi}{a}, \quad m = 1, 2, \dots$$

www.antenna-theory.com

This statement implies that the only functions of x that satisfy the differential equation and the required boundary conditions must be an integer multiple of half-sinusoids within the waveguide.

Combining these results, we can write the solution for F_z as:

$$F_z = A \cos(k_x x) \cos(k_y y) e^{-jk_z z}$$

$$= A \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

In the above, we have combined the remaining nonzero constants c_1 , c_3 , and c_6 into a single constant, A , for simplicity. We have found that only certain distributions (or field configurations) will satisfy the required differential equations and the boundary conditions. Each of these field configurations will be known as a mode. Because we derived the results above for the TE case, the modes will be known as TE_{mn} , where m indicates the number of half-cycle variations within the waveguide for $X(x)$, and n indicates the number of half-cycle variations within the waveguide for $Y(y)$.

Using the field relationships:

$E_x = \frac{-1}{\epsilon} \frac{\partial F_z}{\partial y}$ $E_y = \frac{1}{\epsilon} \frac{\partial F_z}{\partial x}$ $E_z = 0$	$H_x = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 F_z}{\partial x \partial z}$ $H_y = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 F_z}{\partial y \partial z}$ $H_z = -j \frac{1}{\omega \mu \epsilon} \left(\frac{\partial^2}{\partial z^2} + k^2 \right) F_z$
--	---

We can write the allowable field configurations for the TE (transverse electric) modes within a waveguide:

$E_x = A_{mn} \frac{n\pi}{b\epsilon} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$ $E_y = -A_{mn} \frac{m\pi}{a\epsilon} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$ $E_z = 0$	[1]
--	-----

www.antenna-theory.com

$H_x = A_{mn} \frac{m\pi k_z}{a\omega\mu\epsilon} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$ $H_y = A_{mn} \frac{n\pi k_z}{b\omega\mu\epsilon} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$ $H_z = -jA_{mn} \frac{k^2 - k_z^2}{\omega\mu\epsilon} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$	[2]
--	-----

In the above, the constants are written as A_{mn} - this implies that the amplitude for each mode can be independent of the others; however, the field components for a single mode must all be related (that is, E_x and H_y do not have independent coefficients).

Cutoff Frequency (f_c)

At this point in the analysis, we are able to say something intelligent. Recall that the components of the wavenumber must satisfy the relationship:

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \omega^2 \mu \epsilon = \left(\frac{2\pi f}{c} \right)^2 \quad [3]$$

www.antenna-theory.com

Since k_x and k_y are restrained to only take on certain values, we can plug this fact in:

$$\left(\frac{m\pi}{a} \right)^2 + \left(\frac{n\pi}{b} \right)^2 + k_z^2 = k^2 = \left(\frac{2\pi f}{c} \right)^2 \quad [4]$$

An interesting question arises at this point: What is the lowest frequency in which the waveguide will propagate the TE mode?

For propagation to occur, $k_z^2 > 0$. If this is true, then k_z is a real number, so that the field components (equations [1] and [2]) will contain complex exponentials, which represent propagating waves. If on the other hand, $k_z^2 < 0$, then k_z will be an imaginary number, in which case the complex exponential above in equations [1-2] becomes a decaying real exponential. In this case, the fields will not propagate but instead quickly die out within the waveguide. Electromagnetic fields that die off instead of propagate are referred to as evanescent waves.

To find the lowest frequency in which propagation can occur, we set $kz=0$. This is the transition between the cutoff region (evanescent) and the propagation region.

Setting $kz=0$ in equation [4], we obtain:

$$\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 = \left(\frac{2\pi f}{c}\right)^2 \quad [5]$$

If m and n are both zero, then all of the field components in [1-2] become zero, so we cannot have this condition. The lowest value the left hand side of equation [5] can take occurs when m=1 and n=0. The solution to equation [5] when m=1 and n=0, gives the cutoff frequency for this waveguide:

$$f_c = \frac{c}{2a}$$

Any frequency below the cutoff frequency (f_c) will only result in evanescent or decaying modes. The waveguide will not transport energy at these frequencies. In addition, if the waveguide is operating at a frequency just above f_c , then the only mode that is a propagating mode will be the TE₁₀ mode. All other modes will be decaying. Hence, the TE₁₀ mode, since it has the lowest cutoff frequency, is referred to as the dominant mode.

Every mode that can exist within the waveguide has its own cutoff frequency. That is, for a given mode to propagate, the operating frequency must be above the cutoff frequency for that mode. By solving [5] in a more general form, the cutoff frequency for the TE_{mn} mode is given by:

$$f_c^{mn} = \frac{c}{2} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}$$

Although we haven't discussed the TM (transverse magnetic) mode, it will turn out that the dominant TM mode has a higher cutoff frequency than the dominant TE mode.

Determining the fields for the TMz (Transverse Magnetic to the z direction) modes follows a similar procedure to that for the TEz case. To begin, we'll start by discussing the magnetic vector potential, A . This is a non-physical quantity that is often in used antenna theory to simplify the mathematics of Maxwell's Equations.

To understand the magnetic vector potential, note that since the magnetic flux density \mathbf{B} must always be divergenceless:

$$\nabla \cdot \mathbf{B} = 0$$

If a vector quantity is divergenceless, then it can be expressed as the curl of another vector quantity. In math notation, this means that \mathbf{B} can be written as:

$$\mathbf{B} = \nabla \times \mathbf{A}$$

$$\mathbf{H} = \frac{1}{\mu} \nabla \times \mathbf{A}$$

In a source free region, it can be shown that \mathbf{A} must satisfy the wave equation:

$$\nabla^2 \mathbf{A} + \mathbf{k}^2 \mathbf{A} = 0$$

In addition, the TMz fields can be found from the A_z component of the magnetic vector potential, via the following relationships:

$$E_x = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 A_z}{\partial x \partial z}$$

$$E_y = -j \frac{1}{\omega \mu \epsilon} \frac{\partial^2 A_z}{\partial y \partial z}$$

$$E_z = -j \frac{1}{\omega \mu \epsilon} \left(\frac{\partial^2}{\partial z^2} + k^2 \right) A_z$$

$$H_x = \frac{1}{\mu} \frac{\partial A_z}{\partial y}$$

$$H_y = -\frac{1}{\mu} \frac{\partial A_z}{\partial x}$$

$$H_z = 0$$

www.antenna-theory.com

To solve for A_z (and hence determine the E- and H- fields), we follow the same procedure as for the TEz case. That is, we use separation of variables and solve the wave equation for the z-component of \mathbf{A} , then apply boundary conditions that force

the tangential components of the E-fields to be zero on the metallic surfaces.

Performing this procedure, which will not be repeated here, we obtain the solution for A_z :

$$A_z = B_{mn} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z} \quad [1]$$

The corresponding TMz fields for waves propagating in the +z-direction are:

$$E_x = -B_{mn} \frac{k_x k_y}{\omega \mu \epsilon} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

$$E_y = -B_{mn} \frac{k_y k_z}{\omega \mu \epsilon} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

$$E_z = -jB_{mn} \frac{k^2 - k_z^2}{\omega \mu \epsilon} \sin\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

www.antenna-theory.com

$$H_x = B_{mn} \frac{k_y}{\mu} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

$$H_y = -B_{mn} \frac{k_x}{\mu} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

$$H_z = 0$$

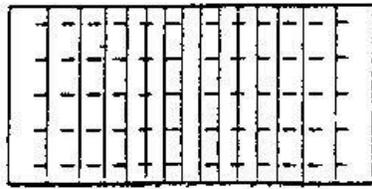
In the above, k is again the wavenumber, and B_{mn} is a constant, which determines the amplitude of the mn mode (a function of how much power is applied to the waveguide at that frequency).

Before discussing the modes, we must note that TM_{0n} and TM_{m0} modes cannot exist; that is, m and n must be at least 1. The reason comes from equation [1] above - if either m or n are zero, then A_z is equal to zero, so all the fields derived must also be zero. Hence, the lowest order mode for the TM case is the TM₁₁ mode.

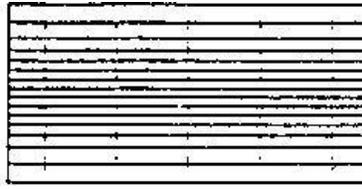
The same procedures can be applied from the TE case to determine the cutoff frequencies for the TM_{mn} mode:

$$f_c^{mn} = \frac{c}{2} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}$$

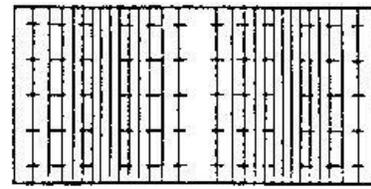
MODES



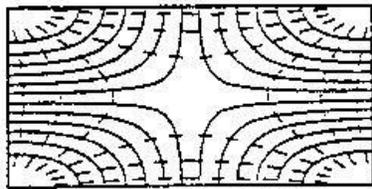
TE₁₀



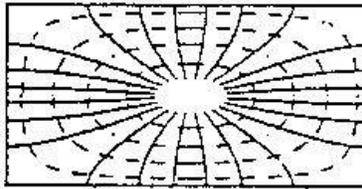
TE₀₁



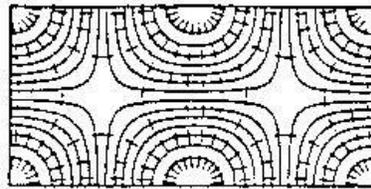
TE₂₀



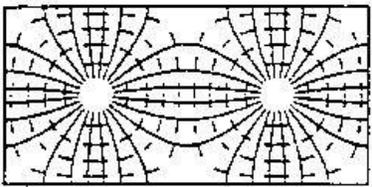
TE₁₁



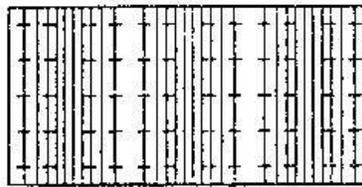
TM₁₁



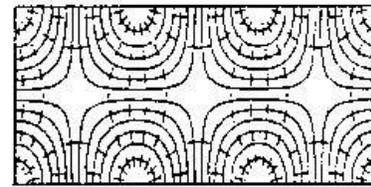
TE₂₁



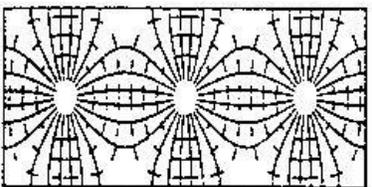
TM₂₁



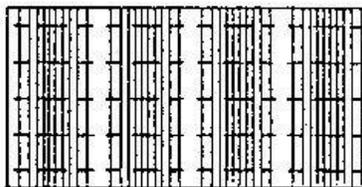
TE₃₀



TE₃₁



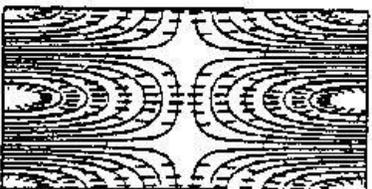
TM₃₁



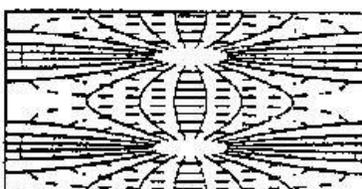
TE₄₀



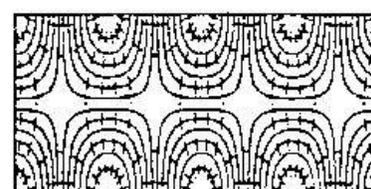
TE₀₂



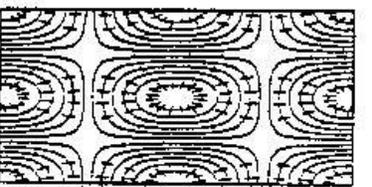
TE₁₂



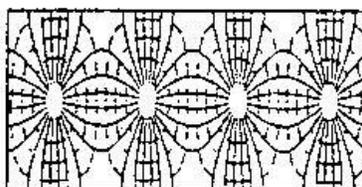
TM₁₂



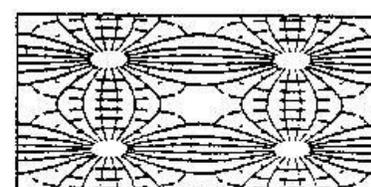
TE₄₁



TE₂₂



TM₄₁



TM₂₂

E —————

H - - - - -

$a/b = 2$

Cylindrical waveguide

Using the complete formulation in the simplest limit possible, global electromagnetic modes are here studied in a large aspect ratio, circular cross-section vacuum cavity equivalent to a cylindrical waveguide.

Classical electrodynamics [43] show that the EM eigenmode spectrum consists of two types of solutions, the transverse electric $TE_{l,m}$ and the transverse magnetic $TM_{l,m}$ polarizations with frequencies depending on l the radial and m the azimuthal mode numbers. These results are reproduced numerically to verify that the wave equations (16) can indeed be solved in the vacuum using standard LFEM and CFEM discretizations, without introducing spurious modes of numerical origin. It is also important to validate the numerical implementation using a simple test case, checking that the numerical solutions converge to the analytical values with rates expected from the order of the approximations.

The cylindrical waveguide is modeled in 2-D, with a circular large aspect ratio equilibrium defined with a minor radius a chosen so as to obtain the analytical eigenmode frequencies in GHz exactly equal to the roots of the Bessel function (table 1).

minor radius	$a[cm]$	4.771345159
aspect ratio	R_0/a	10000
safety factor on axis	q_0	10^3
toroidal wave number	n	0
antenna $I_s(m) = I_0(m)/10^4$	m	$\in \{1, \dots, 4\}$
dielectric tensor	$\vec{\epsilon}$	I
damping	δ_0	$10^{-2} \dots 10^{-8}$
homogeneous mesh	$N = N_s = N_\theta$	4...48

Table 1: Cylindrical waveguide parameters.

As the equilibrium merely produces the geometry and the mesh, the safety factor q_0 does not affect the eigenfrequency spectrum; using a large value for q_0 , it is

however possible to everywhere align $\vec{e}_{||}$ with the axis of the cylinder and separate the components of the TE and the TM polarizations. The complete toroidal wave

equations (16) are then discretized in the large aspect ratio cavity, relying on numerical cancellations to recover the cylindrical limit.

To compute the eigenmode spectrum, an oscillating source current (eq.22) is driven with a small imaginary part in the excitation frequency

$(\omega_r; \omega_i) = 2\pi(1 + i\delta_0)f$. The power relation (eq.48) yields a complex response

function $P^{Pa}(\omega_r; \omega_i)$ which has poles $(\omega_r; 0)$ along the real axis that correspond to the solutions of the discretized wave equations. The eigenfrequencies are calculated by scanning in the complex plane (ω_r, ω_i) with an increment $\Delta\omega_r = 2\pi\Delta f$ and a constant $\Im m(P^{Pa})$

$\omega_i > 0$ chosen so as to resolve the response peaks in . The structure of an

eigenmode is obtained in the limit $\omega_i \ll \omega_r$ when the cavity is resonantly excited at

the maximum of a narrow response peak.

In order to verify that the eigenfrequency spectrum of this cylindrical waveguide is complete and does not contain any spurious "polluting" mode, two broad scans are performed from 10 kHz to 10 GHz with a high resolution in frequency $(\Delta f/f = 0.025, \delta_0 = 0.05)$ and a low resolution in space $(N_r = N_\theta = 8)$ for LFEM,

$N_r = N_\theta = 4$ for CFEM). All the Fourier modes $\{0, \dots, N_\theta/2\}_{I_a(m)}$ representable by the

numerical discretization are excited with azimuthal currents $I_a(m)$ for TE modes,

and axial currents for TM modes. Fig.6 summarizes the result obtained with

LFEM, showing that every mode found numerically could be identified in a one to one correspondence with the analytical result. Modes which have low quantum

numbers (l,m) are, as expected, obtained with a better precision; pushing the resolution to the lowest limit of 2 mesh points per wavelength (m=4), the

deviations become of course important, but the spectrum remains unpolluted (remember fig.3, root b). The same analysis has been repeated with CFEM and leads to results which are much more precise. As an illustration, the eigenmode

$$N_1 = N_2 = 6$$

$H_{1,1}$ has here been calculated on a coarse homogeneous mesh. The eigenfrequency obtained numerically $f_{num} = 5.3323$ GHz is in excellent agreement

with the analytical result $f^{th} = 5.3314$ GHz; fig.5 shows the eigenmode structure in a vector plot of $\vec{E}_\perp \sim \vec{A}_\perp = A_n \vec{e}_n + A_b \vec{e}_t$.

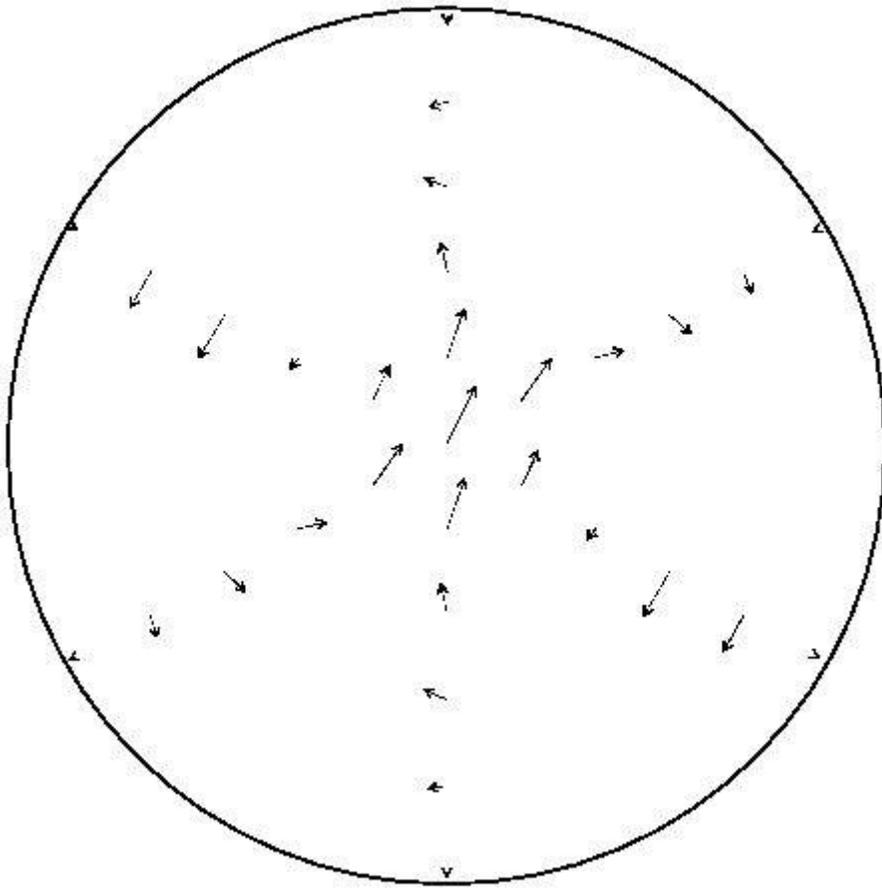


Figure 5: $\text{Re}(A_{\text{perp}})$ for an eigenmode TE_{11} calculated with CFEM.

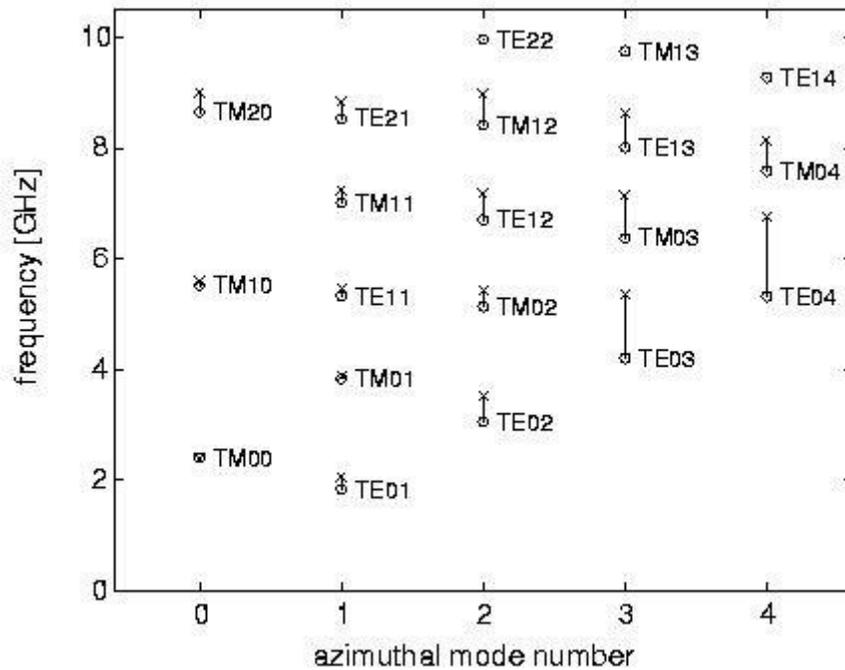


Figure 6: Analytical (circles) and LFEM (x-marks) eigenfrequency spectrum.

A question remained when the boundary conditions were defined in sect.2.2.2: it concerned the implementation of the regularity conditions which is formally not sufficient to forbid a weakly singular ($f(s) \sim \ln(s)$) behavior of the field in the

center of the mesh. Fig.5 shows that the field is regular all over the cylinder radius, suggesting that the singularity is not strong enough to show up using a FEM

discretization on a regular mesh. The only way we have found to observe it, was to strongly accumulate the mesh points towards the center (for example by dividing 5 times the radial mesh interval closest to the axis by two, leading to radial mesh

spacings $\Delta s \sim 1, 2, 4, 8, 16\dots$).

Having verified that the solutions calculated with the wave equations (eq.16) behave in a satisfactory manner, the quality of the LFEM and CFEM discretizations is finally best judged in a convergence study monitoring the precision of the frequency and the gauge as a function of the spatial resolution.

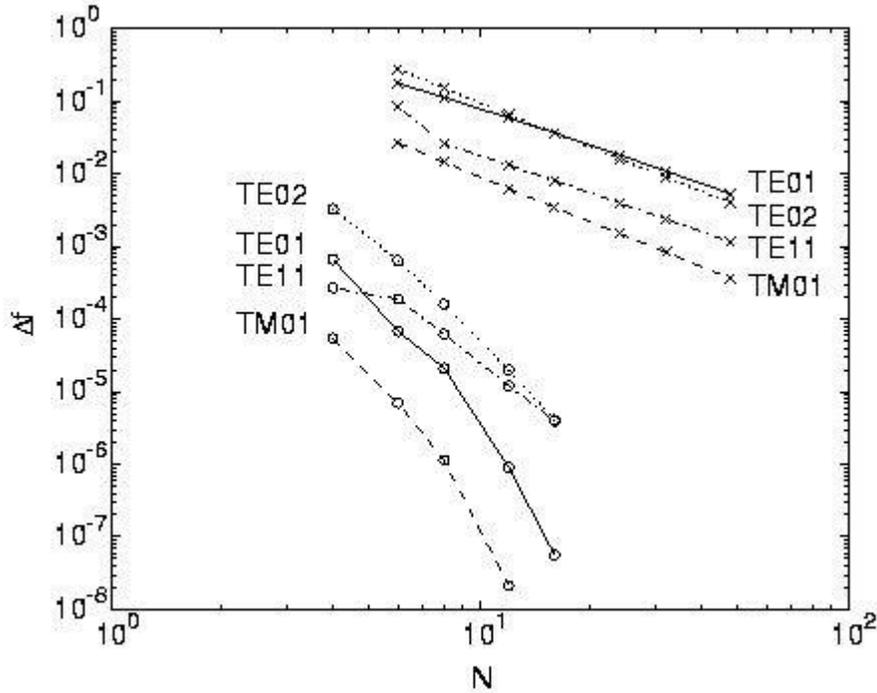


Figure 7: Convergence to the analytical result: relative frequency deviation Δf versus the number of mesh intervals ($N=N_s=N_{\theta}$) for the eigenmodes $\{TE_{01}, TE_{02}, TE_{11}, TM_{00}\}$ using LFEM (x-marks) and CFEM (circles).

Fig.7 shows the convergence of $\Delta f = \frac{(f^{num} - f^{th})}{f^{th}}$ for the eigenmodes TE_{02} , TE_{01} , TE_{11} and TM_{00} , where f^{num} refers to the frequency obtained numerically and f^{th} to the analytical result. Eigenfrequencies converge to the analytical values as N^{-2} using LFEM and almost N^{-5} using CFEM, with an excellent initial precision better than 1% for two mesh points per wavelength.

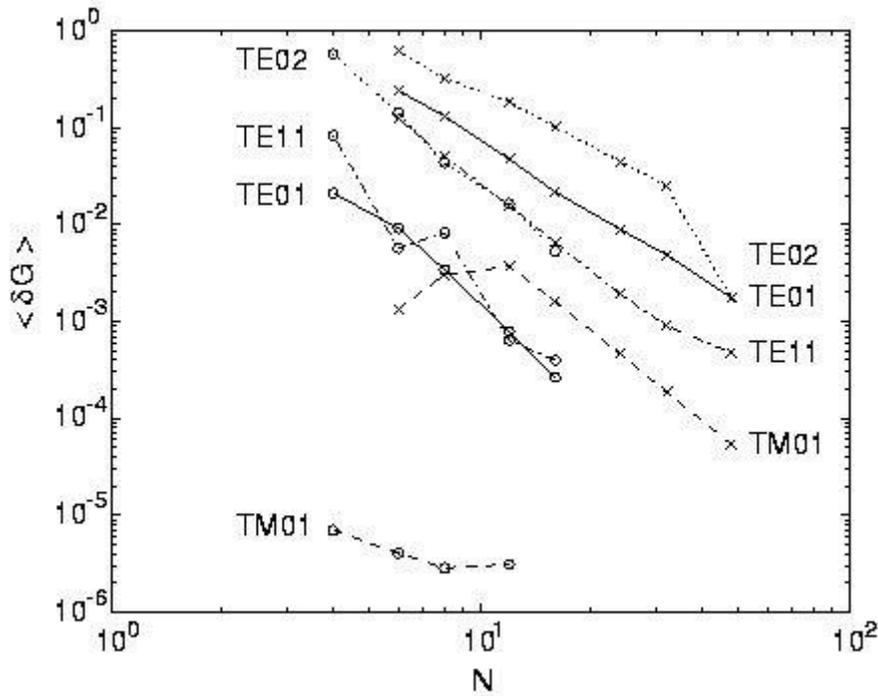


Figure 8: Precision of the gauge versus the number of mesh intervals ($N=N_s=N_{\theta}$) for the eigenmodes TE_01,TE_02,TE_11,TM_00 using LFEM (x-marks) and CFEM (circles).

Convergence is also achieved for the gauge: fig.8 shows that the volume averaged

gauge precision $\langle \delta G \rangle = \left\langle \frac{|\nabla \cdot \vec{A}|}{|\vec{A}|} \right\rangle$ converges to zero as N^{-2} using LFEM, and N^{-4} using CFEM.

To summarize, the calculations performed with the toroidal PENN code used here in the simplest limit possible show that Maxwell's equations (16) solved in a cylindrical cavity produce the complete physical spectrum without introducing numerically produced "polluting" modes. Both, the LFEM and the CFEM discretization schemes yield solutions which are numerically sane and converge to

the analytical value with rates expected from the order of the interpolations.

Boundary conditions

Let us review the general boundary conditions on the field vectors at a surface between medium 1 and medium 2:

$$\mathbf{n} \cdot (\mathbf{D}_1 - \mathbf{D}_2) = \tau, \quad (34)$$

$$\mathbf{n} \wedge (\mathbf{E}_1 - \mathbf{E}_2) = 0, \quad (35)$$

$$\mathbf{n} \cdot (\mathbf{B}_1 - \mathbf{B}_2) = 0, \quad (36)$$

$$\mathbf{n} \wedge (\mathbf{H}_1 - \mathbf{H}_2) = \mathbf{K}, \quad (37)$$

where τ is used for the surface charge density (to avoid confusion with the conductivity), and \mathbf{K} is the surface current density. Here, \mathbf{n} is a unit vector normal to the surface, directed from medium 2 to medium 1. We have seen in Section 4.4

that for normal incidence an electromagnetic wave falls off very rapidly inside the surface of a good conductor. Equation (4.35) implies that in the limit of perfect conductivity ($\sigma \rightarrow \infty$) the tangential component of \mathbf{E} vanishes, whereas that of \mathbf{H} may remain finite. Let us examine the behaviour of the normal components.

Let medium 1 be a good conductor for which $\sigma / \epsilon_0 \omega \gg 1$, whilst medium 2 is a perfect insulator. The surface charge density is related to the currents flowing inside the conductor. In fact, the conservation of charge requires that

$$\mathbf{n} \cdot \mathbf{j} = \frac{\partial \tau}{\partial t} = -i\omega \tau. \quad (38)$$

However, $\mathbf{n} \cdot \mathbf{j} = \mathbf{n} \cdot \sigma \mathbf{E}_1$, so it follows from Eq. (6.1)(a) that

$$\left(1 + \frac{i\omega \epsilon_0 \epsilon_1}{\sigma}\right) \mathbf{n} \cdot \mathbf{E}_1 = \frac{i\omega \epsilon_0 \epsilon_2}{\sigma} \mathbf{n} \cdot \mathbf{E}_2. \quad (39)$$

It is clear that the normal component of \mathbf{E} within the conductor also becomes vanishingly small as the conductivity approaches infinity.

If \mathbf{E} vanishes inside a perfect conductor then the curl of \mathbf{E} also vanishes, and the time rate of change of \mathbf{B} is correspondingly zero. This implies that there are no oscillatory fields whatever inside such a conductor, and that the boundary values of the fields outside are given by

$$\mathbf{n} \cdot \mathbf{D} = -\tau, \quad (40)$$

$$\mathbf{n} \wedge \mathbf{E} = 0, \quad (41)$$

$$\mathbf{n} \cdot \mathbf{B} = 0, \quad (42)$$

$$\mathbf{n} \wedge \mathbf{H} = -\mathbf{K}. \quad (43)$$

Here, \mathbf{n} is a unit normal at the surface of the conductor pointing into the conductor. Thus, the electric field is normal and the magnetic field tangential at the surface of a perfect conductor. For good conductors these boundary conditions yield excellent representations of the geometrical configurations of external fields, but they lead to the neglect of some important features of real fields, such as losses in cavities and signal attenuation in wave guides.

In order to estimate such losses it is useful to see how the tangential and normal fields compare when σ is large but finite. Equations (4.5) and (4.34) yield

$$\mathbf{H} = \frac{1+i}{\sqrt{2}} \sqrt{\frac{\sigma}{\mu_0 \omega}} \mathbf{n} \wedge \mathbf{E} \quad (44)$$

at the surface of a conductor (provided that the wave propagates into the conductor). Let us assume, without obtaining a complete solution, that a wave with

\mathbf{H} very nearly tangential and \mathbf{E} very nearly normal is propagated along the surface of the metal. According to the Faraday-Maxwell equation

$$|H_{||}| \simeq \frac{k}{\mu_0 \omega} |E_{\perp}| \quad (45)$$

just outside the surface, where k is the component of the propagation vector along the surface. However, Eq. (6.5) implies that a tangential component of \mathbf{H} is accompanied by a small tangential component of \mathbf{E} . By comparing these two expressions, we obtain

$$\frac{|E_{||}|}{|E_{\perp}|} \simeq k \sqrt{\frac{2}{\mu_0 \omega \sigma}} = \frac{d}{\lambda}, \quad (46)$$

where d is the skin depth (see Eq. (4.36)) and $\lambda \equiv 1/k$. It is clear that the ratio of the tangential component of \mathbf{E} to its normal component is of order the skin depth divided by the wavelength. It is readily demonstrated that the ratio of the normal component of \mathbf{H} to its tangential component is of this same magnitude. Thus, we can see that in the limit of high conductivity, which means vanishing skin depth, no fields penetrate the conductor, and the boundary conditions are those given by Eqs. (6.4). Let us investigate the solution of the homogeneous wave equation subject to such boundary conditions.

Cavities with rectangular boundaries

Consider a vacuum region totally enclosed by rectangular conducting walls. In this

case, all of the field components satisfy the wave equation

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0, \quad (47)$$

where ψ represents any component of \mathbf{E} or \mathbf{H} . The boundary conditions (6.4)

require that the electric field is normal to the walls at the boundary whereas the magnetic field is tangential. If a , b , and c are the dimensions of the cavity, then it is readily verified that the electric field components are

$$E_x = E_1 \cos(k_1 x) \sin(k_2 y) \sin(k_3 z) e^{-i\omega t}, \quad (48)$$

$$E_y = E_2 \sin(k_1 x) \cos(k_2 y) \sin(k_3 z) e^{-i\omega t}, \quad (49)$$

$$E_z = E_3 \sin(k_1 x) \sin(k_2 y) \cos(k_3 z) e^{-i\omega t}, \quad (50)$$

where

$$k_1 = \frac{l\pi}{a}, \quad (51)$$

$$k_2 = \frac{m\pi}{b}, \quad (52)$$

$$k_3 = \frac{n\pi}{c}, \quad (53)$$

with l , m , n integers. The allowed frequencies are given by

$$\frac{\omega^2}{c^2} = \pi^2 \left(\frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{n^2}{c^2} \right). \quad (54)$$

It is clear from Eq. (6.9) that at least two of the integers l , m , n must be different from zero in order to have non-vanishing fields. The magnetic fields obtained by

the use of $\nabla \wedge \mathbf{E} = i\omega \mathbf{B}$ automatically satisfy the appropriate boundary conditions, and are in phase quadrature with the electric fields. Thus, the sum of the total electric and magnetic energies within the cavity is constant, although the two terms oscillate separately.

The amplitudes of the electric field components are not independent, but are related by the divergence condition $\nabla \cdot \mathbf{E} = 0$, which yields

$$k_1 E_1 + k_2 E_2 + k_3 E_3 = 0. \quad (55)$$

There are, in general, two linearly independent vectors \mathbf{E} that satisfy this condition, corresponding to two polarizations. (The exception is the case that one of the integers l, m, n is zero, in which case \mathbf{E} is fixed in direction.) Each vector is accompanied by a magnetic field at right angles. The fields corresponding to a given set of integers l, m, n constitute a particular mode of vibration of the cavity. It is evident from standard Fourier theory that the different modes are orthogonal (i.e., they are normal modes) and that they form a complete set. In other words, any general electric and magnetic fields which satisfy the boundary conditions (6.4) can be unambiguously decomposed into some linear combination of all of the various possible normal modes of the cavity. Since each normal mode oscillates at a specific frequency it is clear that if we are given the electric and magnetic fields inside the cavity at time $t = 0$ then the subsequent behaviour of the fields is uniquely determined for all time.

The conducting walls gradually absorb energy from the cavity, due to their finite resistivity, at a rate which can easily be calculated. For finite σ the small tangential component of \mathbf{E} at the walls can be estimated using Eq. (6.5):

$$\mathbf{E}_{||} = \frac{1-i}{\sqrt{2}} \sqrt{\frac{\mu_0 \omega}{\sigma}} \mathbf{H}_{||} \wedge \mathbf{n}. \quad (56)$$

Now, the tangential component of \mathbf{H} at the walls is slightly different from that given by the ideal solution. However, this is a small effect and can be neglected to leading order in σ^{-1} . The time averaged energy flux into the walls is given by

$$\overline{N} = \frac{1}{2} \text{Re}(\mathbf{E}_{||} \wedge \mathbf{H}_{||}) = \frac{1}{2} \sqrt{\frac{\mu_0 \omega}{2\sigma}} H_{||0}^2 \mathbf{n} = \frac{H_{||0}^2}{2\sigma d} \mathbf{n}, \quad (57)$$

where $H_{||0}$ is the peak value of the tangential magnetic field at the walls predicted by the ideal solution. According to the boundary condition (6.4)(d), $H_{||0}$ is equal to

K_0 the peak value of the surface current density. It is helpful to define a surface resistance,

$$\overline{N} = \overline{K^2} R_s \mathbf{n}, \quad (58)$$

where

$$R_s = \frac{1}{\sigma d}. \quad (59)$$

This approach makes it clear that the dissipation of energy is due to ohmic heating in a thin layer, whose thickness is of order the skin depth, on the surface of the conducting walls.

The quality factor of a resonant cavity

The quality factor Q of a resonant cavity is defined

$$Q = 2\pi \frac{\text{energy stored in cavity}}{\text{energy lost per cycle to walls}} \quad (60)$$

For a specific normal mode of the cavity this quantity is independent of the mode amplitude. By conservation of energy the power dissipated in ohmic losses is minus the rate of change of the stored energy U . We can write a differential equation for the behaviour of U as a function of time:

$$\frac{dU}{dt} = -\frac{\omega_0}{Q} U, \quad (61)$$

where ω_0 is the oscillation frequency of the normal mode in question. The solution to the above equation is

$$U(t) = U(0) e^{-\omega_0 t / Q}. \quad (62)$$

This time dependence of the stored energy suggests that the oscillations of the fields in the cavity are damped as follows:

$$E(t) = E_0 e^{-\omega_0 t / 2Q} e^{-i(\omega_0 + \Delta\omega)t}, \quad (63)$$

where we have allowed for a shift $\Delta\omega$ of the resonant frequency as well as the damping. A damped oscillation such as this does not consist of a pure frequency.

$$\omega = \omega_0 + \Delta\omega$$

Instead, it is made up of a superposition of frequencies around .

Standard Fourier analysis yields

$$E(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t} d\omega, \quad (64)$$

where

$$E(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} E_0 e^{-\omega_0 t / 2Q} e^{i(\omega - \omega_0 - \Delta\omega)t} dt. \quad (65)$$

It follows that

$$|E(\omega)|^2 \propto \frac{1}{(\omega - \omega_0 - \Delta\omega)^2 + (\omega_0 / 2Q)^2}. \quad (66)$$

The resonance shape has a full width at half-maximum equal to ω_0 / Q . For a constant input voltage, the energy of oscillation in the cavity as a function of frequency follows the resonance curve in the neighbourhood of a particular resonant frequency. It can be seen that the ohmic losses, which determine for a particular mode, also determine the maximum amplitude of the oscillation when the resonance condition is exactly satisfied, as well as the width of the resonance (i.e., how far off the resonant frequency the system can be driven and still yield a significant oscillation amplitude).

Cylindrical cavities

Let us apply the methods of the previous section to the TM modes of a right circular cylinder of radius a . We can write

$$E_z = A_z f(r, \varphi) \cos(k_3 z) e^{-i\omega t},$$

where $f(r, \varphi)$ satisfies the equation

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial f}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 f}{\partial \varphi^2} + k_1^2 f = 0,$$

and (r, φ, z) are cylindrical polar coordinates. Let

$$f(r, \varphi) = g(r) e^{im\varphi}.$$

It follows that

$$\frac{1}{r} \frac{d}{dr} \left(r \frac{dg}{dr} \right) + \left(k_1^2 - \frac{m^2}{r^2} \right) g = 0,$$

or

$$z^2 \frac{d^2 g}{dz^2} + z \frac{dg}{dz} + (z^2 - m^2) g = 0,$$

where $z = k_1 r$. The above equation is known as Bessel's equation. The two linearly independent solutions of Bessel's equation are denoted $J_m(z)$ and $Y_m(z)$. In the

limit $|z| \ll 1$ these solutions behave as z^m and z^{-m} , respectively, to lowest order .

More exactly¹⁶

$$J_m(z) = \left(\frac{z}{2}\right)^m \sum_{k=0}^{\infty} \frac{(-z^2/4)^k}{k!(m+k)!},$$

$$Y_m(z) = -\frac{(z/2)^{-m}}{\pi} \sum_{k=0}^{m-1} \frac{(m-k-1)!(z^2/4)^k}{k!} + \frac{1}{\pi} \ln(z/2) J_m(z) - \frac{(z/2)^m}{\pi} \sum_{k=0}^{\infty} [\psi(k+1) + \psi(m+k+1)] \frac{(-z^2/4)^k}{k!(m+k)!}$$

for $|z| \ll 1$, where

$$\psi(1) = -\gamma,$$

$$\psi(n) = -\gamma + \sum_{k=1}^{n-1} k^{-1},$$

and $\gamma = \sum_{k=1}^{\infty} k^{-1} = 0.57722$ is Euler's constant. Clearly, the J_m are well behaved in

the limit $|z| \rightarrow 0$, whereas the Y_m are badly behaved.

The asymptotic behaviour of both solutions at large $|z|$ is

$$J_m(z) = \sqrt{\frac{2}{\pi z}} \cos(z - m\pi/2 - \pi/4) + O(1/z),$$

$$Y_m(z) = \sqrt{\frac{2}{\pi z}} \sin(z - m\pi/2 - \pi/4) + O(1/z).$$

Thus, for $|z| \gg 1$ the solutions take the form of gradually decaying oscillations

$$J_0(z) \quad Y_0(z)$$

which are in phase quadrature. The behaviour of $J_0(z)$ and $Y_0(z)$ is shown in Fig. 21.

$$Y_0(z)$$

Figure 21: The Bessel functions $J_0(z)$ (solid line) and $Y_0(z)$ (dotted line)

Since the axis $r = 0$ is included in the cavity the radial eigenfunction must be

$$Y_m(k_r r)$$

regular at the origin. This immediately rules out the $Y_m(k_r r)$ solutions. Thus, the

most general solution for a TM mode is

$$E_z = A J_m(k_r r) e^{im\phi} \cos(k_z z) e^{-i\omega t}.$$

The k_r are the eigenvalues of k_r , and are determined by the solutions of

$$J_m(k_r a) = 0.$$

The above constraint ensures that the tangential electric field is zero on the

conducting walls surrounding the cavity ($r = a$).

The most general solution for a TE mode is

$$H_z = A J_m(k_r r) e^{i m \phi} \sin(k_z z) e^{-i \omega t}.$$

In this case, the k_r are determined by the solution of

$$J'_m(k_r a) = 0,$$

where $'$ denotes differentiation with respect to the argument. The above constraint ensures that the normal magnetic field is zero on the conducting walls surrounding the cavity. The oscillation frequency of both the TM and TE modes is given by

$$\frac{\omega^2}{c^2} = k^2 = k_r^2 + \frac{n^2 \pi^2}{L^2}.$$

If l is the ordinal number of a zero of a particular Bessel function of order m (increases with increasing values of the argument), then each mode is characterized

by three integers, m , l , and n , as in the rectangular case. The l th zero of J'_m is conventionally denoted $j'_{m,l}$ [so, $J'_m(j'_{m,l}) = 0$].

Likewise, the n th zero of J_n is denoted $j_{n,l}$. Table 2 shows the first few zeros of J_0 , J'_0 , J_1 , and J'_1 . It is clear that for fixed n and m the lowest frequency mode (i.e., the mode with the lowest value of k_r) is a TE mode. The mode with the next highest frequency is also a TE mode. The next highest frequency mode is a TM mode, and so on.

Table 2: The first few values of $j_{0,l}$, $j_{1,l}$, $j'_{0,l}$, and $j'_{1,l}$

l	$j_{0,l}$	$j_{1,l}$	$j'_{0,l}$	$j'_{1,l}$
1	2.4048	3.8317	0.0000	1.8412
2	5.5201	7.0156	3.8317	5.3314
3	8.6537	10.173	7.0156	8.5363
4	11.792	13.324	10.173	11.706

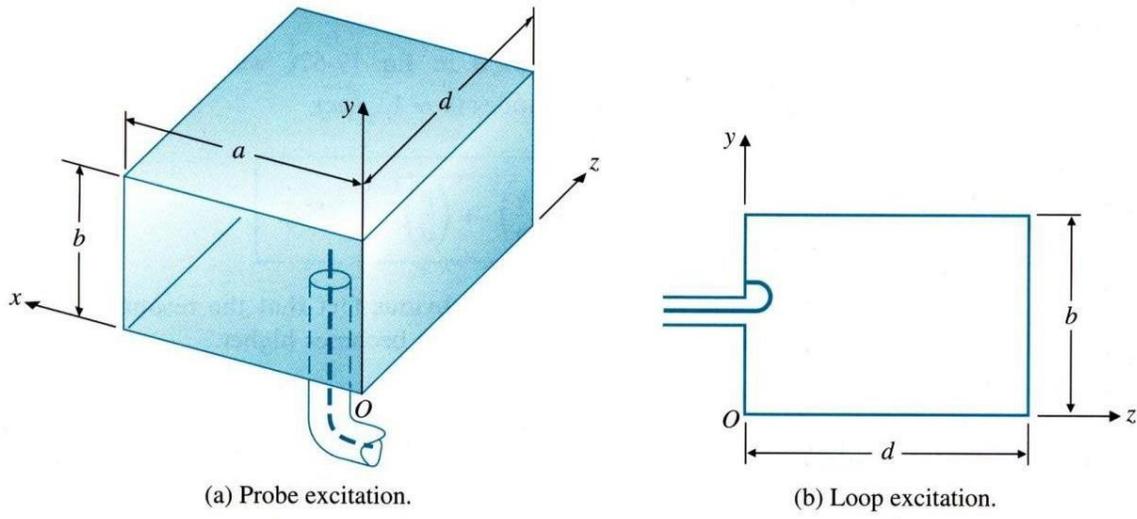
- Cavity resonators are enclosed metal boxes.

Electromagnetic fields are confined inside the boxes. Radiation and high-resistance effects are eliminated, resulting in a very high Q (quality factor)

A rectangular waveguide with both ends ($z=0$ and $z=d$) closed by a conducting wall

(Figure 9-8) : multiple reflections and standing waves

FIGURE 9-8 Excitation of cavity modes by a coaxial line.



Resonant frequency of rectangular cavity resonator

$$f_{mnp} = \frac{u}{2} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2 + \left(\frac{p}{d}\right)^2} \quad (\text{Hz}). \quad (9-103)$$

Degenerate modes : different modes having the same resonant frequency.

Dominant mode : the mode with the lowest resonant frequency for a given cavity size.

A resonator is a device or system that exhibits resonance or resonant behavior, that is, it naturally oscillates at some frequencies, called its resonant frequencies, with greater amplitude than at others. The oscillations in a resonator can be either electromagnetic or mechanical (including acoustic). Resonators are used to either generate waves of specific frequencies or to select specific frequencies from a signal. Musical instruments use acoustic resonators that produce sound waves of specific tones.

A cavity resonator, usually used in reference to electromagnetic resonators, is one in which waves exist in a hollow space inside the device. Acoustic cavity resonators, in which sound is produced by air vibrating in a cavity with one opening, are known as Helmholtz resonators.

Cavity resonators

A cavity resonator is a hollow conductor blocked at both ends and along which an electromagnetic wave can be supported. It can be viewed as a waveguide short-circuited at both ends (see Microwave cavity).

The cavity has interior surfaces which reflect a wave of a specific frequency. When a wave that is resonant with the cavity enters, it bounces back and forth within the cavity, with low loss (see standing wave). As more wave energy enters the cavity,

it combines with and reinforces the standing wave, increasing its intensity.

The cavity magnetron is a vacuum tube with a filament in the center of an evacuated, lobed, circular cavity resonator. A perpendicular magnetic field is

imposed by a permanent magnet. The magnetic field causes the electrons, attracted

to the (relatively) positive outer part of the chamber, to spiral outward in a

circular

path rather than moving directly to this anode. Spaced about the rim of the

chamber are cylindrical cavities. The cavities are open along their length and so

connect the common cavity space. As electrons sweep past these openings they

induce a resonant high frequency radio field in the cavity, which in turn causes the

electrons to bunch into groups. A portion of this field is extracted with a short

antenna that is connected to a waveguide (a metal tube usually of rectangular cross

section). The waveguide directs the extracted RF energy to the load, which may be

a cooking chamber in a microwave oven or a high gain antenna in the case of radar.

The klystron, tube waveguide, is a beam tube including at least two

apertured

cavity resonators. The beam of charged particles passes through the apertures of

the resonators, often tunable wave reflection grids, in succession. A collector

electrode is provided to intercept the beam after passing through the resonators.

The first resonator causes bunching of the particles passing through it. The

bunched particles travel in a field-free region where further bunching occurs, then

the bunched particles enter the second resonator giving up their energy to excite it

into oscillations. It is a particle accelerator that works in conjunction with a

specifically tuned cavity by the configuration of the structures. On the beamline of

an accelerator system, there are specific sections that are cavity resonators for RF.

The reflex klystron is a klystron utilizing only a single apertured cavity resonator through which the beam of charged particles passes, first in one direction.

A

repeller electrode is provided to repel (or redirect) the beam after passage through

the resonator back through the resonator in the other direction and in proper phase

to reinforce the oscillations set up in the resonator.

In a laser, light is amplified in a cavity resonator which is usually composed of two

or more mirrors. Thus an optical cavity, also known as a resonator, is a cavity

with

walls which reflect electromagnetic waves (light). This will allow standing wave

modes to exist with little loss outside the cavity.