TRANSMISSION LINES AND WAVEGUIDES

L T P C
3 0 0 3

OBJECTIVES:

- To become familiar with propagation of signals through lines.
- To know about various line parameters by conventional and graphical methods.
- To under the need for impedance matching and different impedance matching techniques.
- To understand the design of different types of filters, equalizer and attenuators.

INTENDED OUTCOMES:

- Students have complete exposure to basics and Fundamentals of transmission lines and networks.
- Students will have ability to design filters, equalizers and attenuator.

UNIT I   TRANSMISSION LINE THEORY   9

General theory of Transmission lines – the transmission line– general solution–The infinite line–Wavelength, velocity of propagation– Waveform distortion– the distortion lessline- Loading and different methods of loading– Line not terminated in Z0 –Reflection coefficient–Calculation of current, voltage, power delivered and efficiency of transmission–Input and transfer impedance-Open and short-circuited lines–reflection factor and reflection loss.

UNIT II   HIGH FREQUENCY TRANSMISSION LINES   9

Transmission line equations at radio frequencies–Line of Zero dissipation–Voltage   and current on the dissipation less line, StandingWaves, Nodes Standing Wave Ratio–Input impedance of the dissipation less line - Open and short-circuited lines – Powerand impedance measurement on lines – Reflection losses– Measurement of VSWR and wavelength.

UNIT III   IMPEDANCE MATCHING IN HIGH FREQUENCY LINES   9

Impedance matching: Quarter wave transformer – Impedance matching bystubs–Single stuband double stubmatching– Smith chart – Solutions of problems using Smith chart – Single and double stubmatching using Smithc hart.

UNIT IV   PASSIVE FILTERS   9

Characteristic impedance of symmetrical networks–filter fundamentals. Design of filters: Constant K, LowPass, High Pass, Band Pass, Band Elimination, m-derived sections and composite.

UNIT V   ATTENUATORS AND EQUALIZERS   9

Attenuators: Lattice Attenuators, Bridged– T attenuator, L-Type Attenuator. Equalizers: Inverse network, series, fullseries, shunt, fullshunt, constant resistanceT, constant resistance, constant resistance lattice and bridged T network.

Total Hours: 45

TEXTBOOKS:

S.NO.	Author(s)Name	Titleof thebook	Publisher	Year of publication
1	Ryder.J. D	Networks, Linesand Fields	PHI, NewDelhi.	2009
2	E.G. Jordan&balmain	ElectromagneticWaves AndRadiatingSystems, 2ndEdition,	Prentice-Hallof India,	2000

REFERENCES:

S.NO.	Author(s)Name	Titleof thebook	Publisher	Year of publication
1	UmeshSinha	TransmissionLinesand Network	SatyaPrakashan PublishingCompany, NewDelhi	2012
2	Sudhakar.A, ShyammohanS Palli	Circuits and Networks AnalysisandSynthesis 4th Edition	TataMcGrawHill,	2010

UNIT -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 ≈ 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.

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}:

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

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:

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

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

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

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 calcuiate 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 $\sqrt{(1000)}\ \mathrm{V} \approx 31.62\ \mathrm{V}$ to $1\ \mathrm{V}$ in decibels, use the formula

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

Notice that $(31.62\,\mathrm{V}/1\,\mathrm{V})^2 \approx 1\,\mathrm{kW}/1\,\mathrm{W}$, illustrating the consequence from the definitions above that G_{dB} has the same value, $30\ \mathrm{dB}$, 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

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

- 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 $\log(A \times B \times C \times ...) = \log(A) + \log(B) + \log(C) + ...$
- 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 µW (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. $X_{dBm} = X_{dBW} + 30$.

dBW

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

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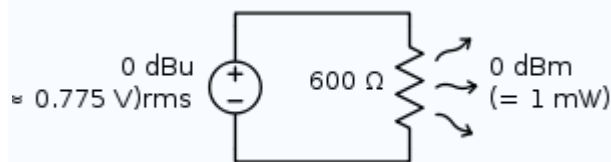
A schematic showing the relationship between dBu (the voltage source) and dBm (the power dissipated as heat by the 600 Ω resistor)

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

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.

dBµV or dBuV

dB(1 µV$_{RMS}$) — voltage relative to 1 microvolt. Widely used in television and aerial amplifier specifications. 60 dBµV = 0 dBmV.

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-pi 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

$V_1 \overset{\text{def}}{=}$ Input voltage
$V_2 \overset{\text{def}}{=}$ Output voltage
$I_1 \overset{\text{def}}{=}$ Input current
$I_2 \overset{\text{def}}{=}$ 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 approach is replaced by 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.

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

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

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):

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

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

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

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

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 sinusiodal 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;

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

The reason for the use of base e is also now made clear. The imaginary phase constant, iβ, 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 (tanδ) of the material, which depends inversely on the wavelength of the signal and is directly proportional to the frequency.

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

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.

1. 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.

Cascaded networks
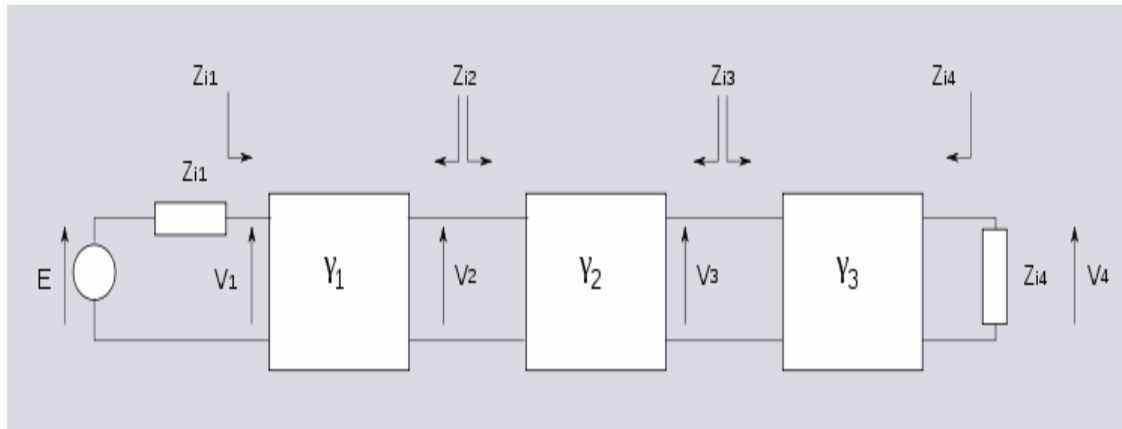
Three networks with arbitrary propagation constants and impedances connected in cascade. The Z_i terms represent image impedance and it is assumed that connections are between matching image impedances.

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,

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

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.

Low Pass

High Pass

Band Pass

Band Reject
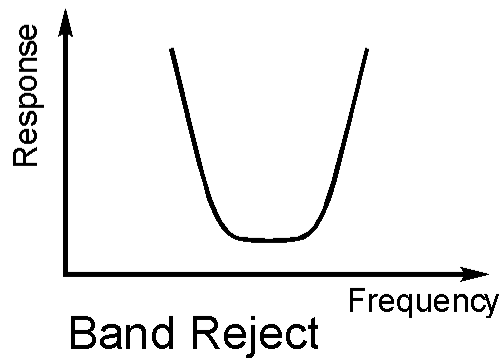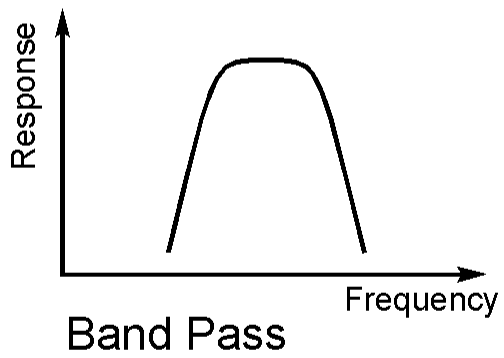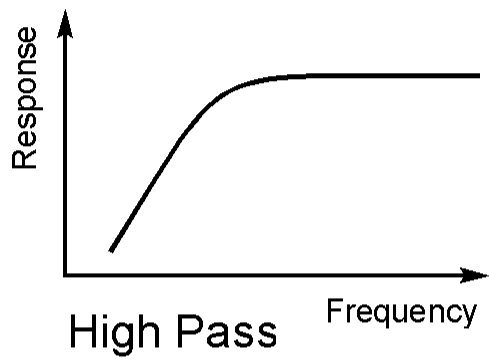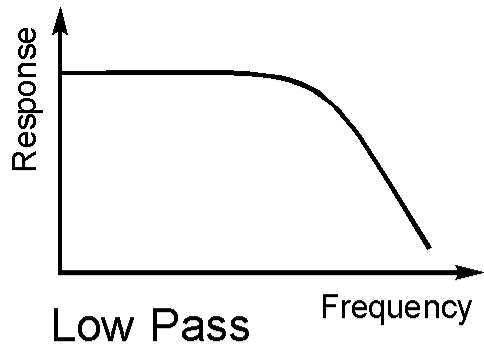
filter frequencies

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.

2. 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 – in electronic filter topology, "L" refers to the specific filter shape which resembles inverted letter "L".
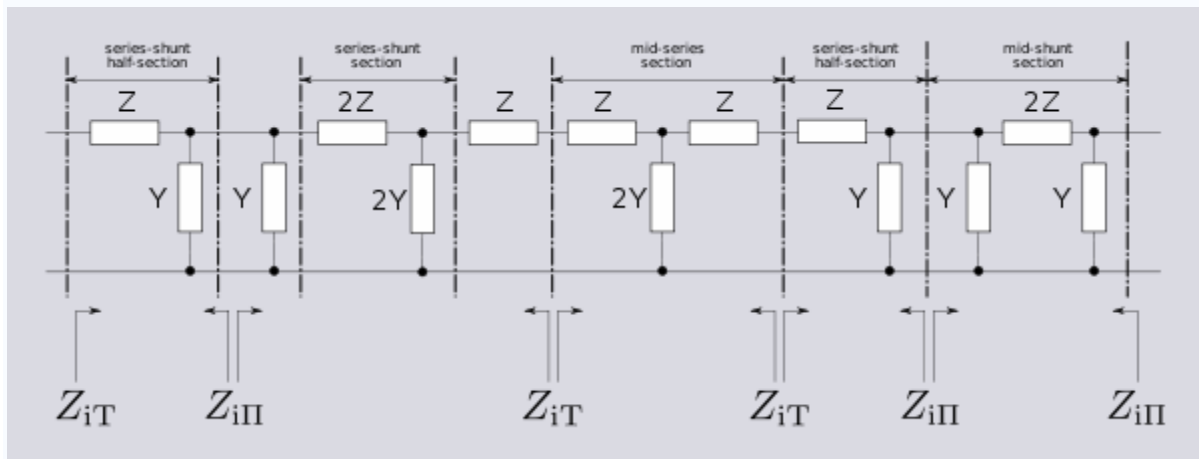
The sections of the hypothetical infinite filter are made of series elements having impedance 2Z and shunt elements with admittance 2Y. The factor of two is introduced for mathematical convenience, since it is usual to work in terms of half-

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 "Π" topology. Half of such a "T" or "Π" 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 variant ways of using a half-section.

Derivation
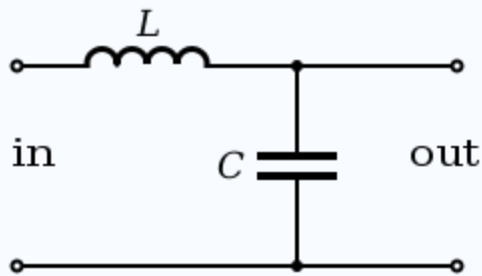
Constant k low-pass filter half section. Here inductance L is equal Ck^2
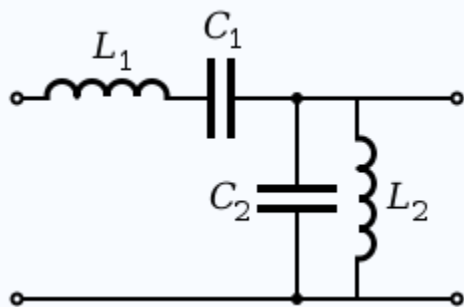
Constant   k   band-pass   filter   half   section.
$L_1 = C_2k^2$ and $L_2 = C_1k^2$
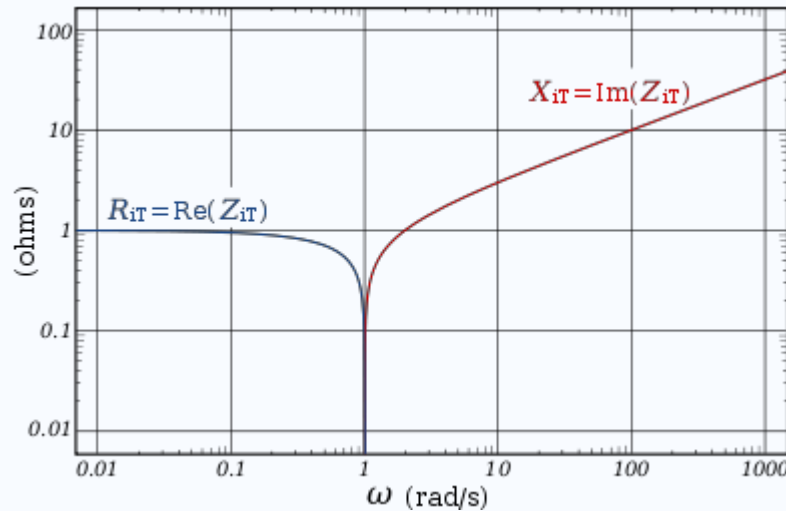
Image impedance Z_{iT} of a constant k prototype low-pass filter is plotted vs. frequency ω. The impedance is purely resistive (real) below $ω_c$, and purely reactive (imaginary) above $ω_c$.

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 ω = 0 in the case of low-pass filters.[7] For example, the pictured low-pass half-section has

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

Elements 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,

$$\lim \; Z_i = k$$

Image impedance

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

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

and

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

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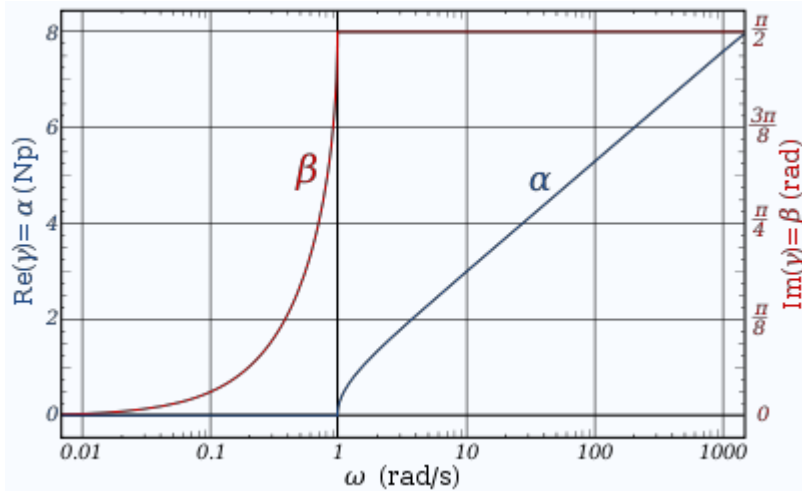
The transfer function of a constant k prototype low-pass filter for a single half-section showing attenuation in nepers and phase change in radians.

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]

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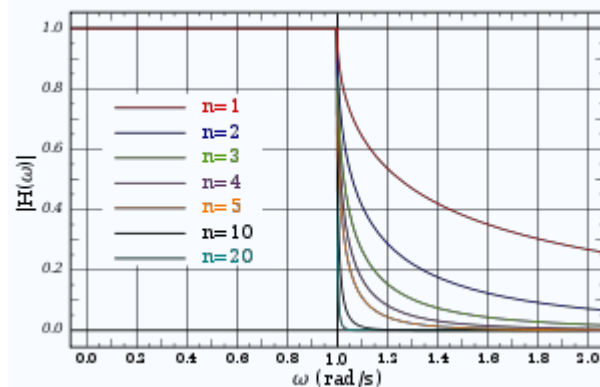
Gain response, $H(\omega)$ for a chain of n low-pass constant-k filter half-sections.

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]

3. 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-types 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
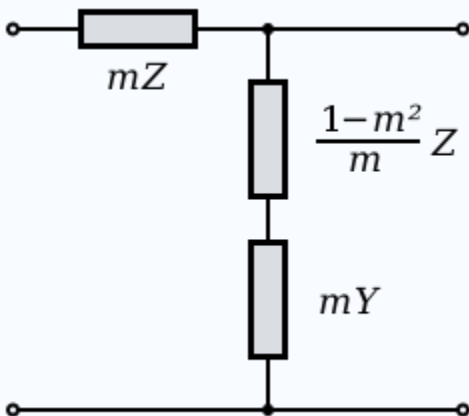
m-derived series general filter half section.
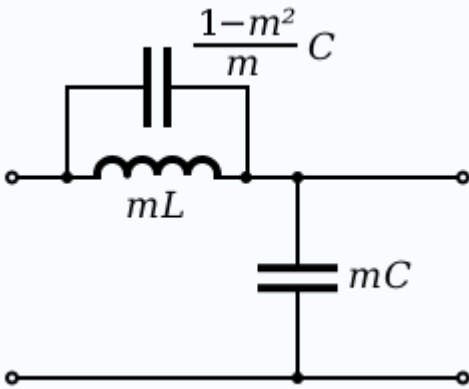
m-derived shunt low-pass filter half section.

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

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

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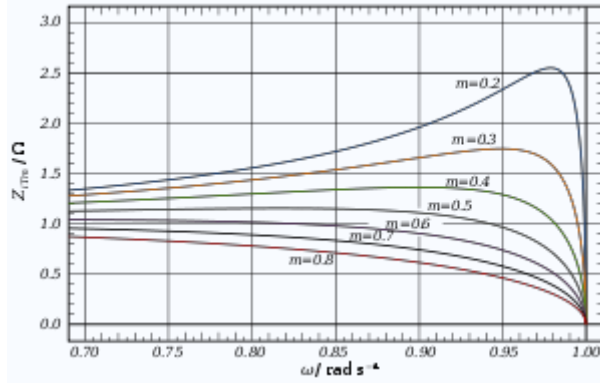
m-derived prototype shunt low-pass filter Z_{iTm} image impedance for various values of m. Values below cut-off frequency only shown for clarity.

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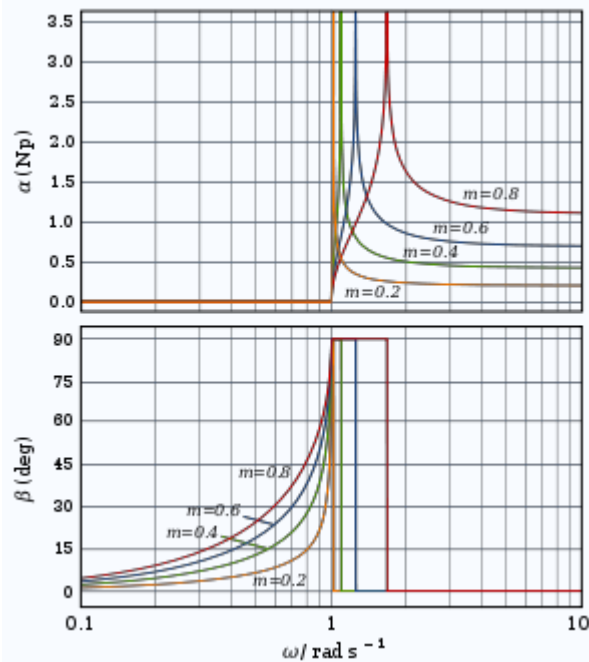
m-Derived low-pass filter transfer function for a single half-section

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 ω_c = 1 rad/s and a nominal impedance R_0 = 1 Ω. 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]

4. 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.

frequency power back toward the source, causing a condition called standing waves. This acts as a bottleneck, reducing the amount of power reaching the destination end of the cable. To hold the shield at a uniform distance from the central conductor, the space between the two is filled with a semirigid plastic dielectric. Manufacturers specify a minimum bend radius[2] to prevent kinks that would cause reflections. The connectors used with coax are designed to hold the correct spacing through the body of the connector.

Each type of coaxial cable has a characteristic impedance depending on its dimensions and materials used, which is the ratio of the voltage to the current in the cable. In order to prevent reflections at the destination end of the cable from causing standing waves, any equipment the cable is attached to must present an impedance equal to the characteristic impedance (called 'matching'). Thus the equipment "appears" electrically similar to a continuation of the cable, preventing reflections. Common values of characteristic impedance for coaxial cable are 50 and 75 ohms.

<div align="center">

UNIT II
TRANSMISSION LINE PARAMETERS

</div>

1. INTRODUCTION

1. THE SYMMETRICAL T NETWORK
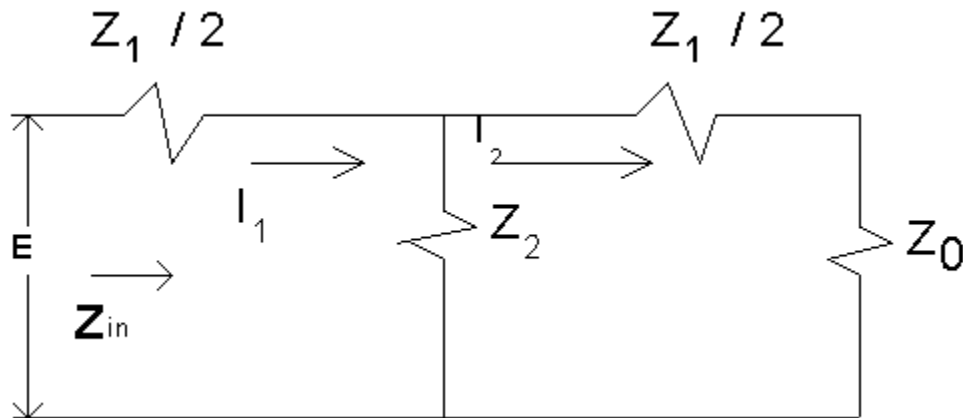
Fig. 1

The value of Z_O (image impedance) for a symmetrical network can be easily determined. For the symmetrical T network of Fig. 1, terminated in its image impedance Z_O, and if $Z_1 = Z_2 = Z_T$ then from many textbooks:

(2.1)
$$Z_{0_T} = \sqrt{Z_1 Z_2\left(1+\frac{Z_1}{4Z_2}\right)}$$

(2.2)
$$Z_{0_T} = \frac{Z_T(\alpha-\beta)}{2} = \frac{Z_T\sqrt{5}}{2} = Z_T\cosh(\ln\alpha)$$

Under Z_O termination, input and output voltage and current are:

(2.3)
$$\frac{I_1}{I_2} = \frac{V_1}{V_2} = e^{\gamma}$$

If there are n such terminated sections then the input and output voltages and currents, under Z_O terminations are:

$$\frac{I_1}{I_2} = \frac{V_1}{V_2} = e^{n\gamma} = \alpha^{2n}$$

$$\frac{I_S}{I_R} = \frac{V_S}{V_R} = e^{\eta l} = \alpha^{2n}$$

Where γ is the propagation constant for one T section., e^{γ} can be evaluated as:

$$e^\gamma = 1 + \frac{Z_1}{2Z_2} + \sqrt{(\frac{Z_1}{2Z_2})^2 + \frac{Z_1}{Z_2}} \qquad e^\gamma = 1 + \frac{Z_1}{2Z_2} + \sqrt{(\frac{Z_1}{2Z_2})^2 + \frac{Z_1}{Z_2}}$$

General solution of the transmission line:

   It is used to find the voltage and current at any points on the transmission line. Transmission lines behave very oddly at high frequencies. In traditional (low-frequency) circuit theory, wires connect devices, but have zero resistance. There is no phase delay across wires; and a short-circuited line always yields zero resistance.

For high-frequency transmission lines, things behave quite differently. For instance, short-circuits can actually have an infinite impedance; open-circuits can behave like short-circuited wires. The impedance of some load (ZL=XL+jYL) can be transformed at the terminals of the transmission line to an impedance much different than ZL. The goal of this tutorial is to understand transmission lines and the reasons for their odd effects.

Let's start by examining a diagram. A sinusoidal voltage source with associated impedance ZS is attached to a load ZL (which could be an antenna or some other device - in the circuit diagram we simply view it as an impedance called a load). The load and the source are connected via a transmission line of length L:
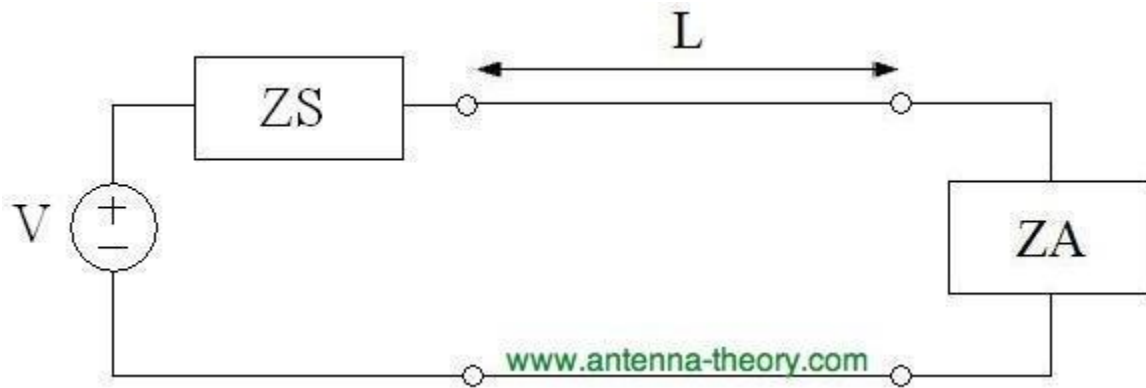
In traditional low-frequency circuit analysis, the transmission line would not matter. As a result, the current that flows in the circuit would simply be:

$$I = \frac{V}{Z} \rightarrow I = \frac{V}{ZS + ZA}$$

However, in the high frequency case, the length L of the transmission line can significantly affect the results. To determine the current that flows in the circuit, we would need to know what the input impedance is, Zin, viewed from the terminals of the transmission line:
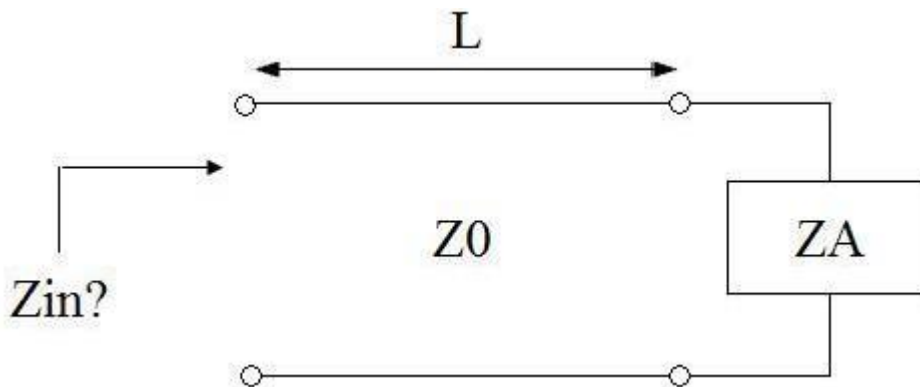

The resultant current that flows will simply be:

$$I = \frac{V}{Z} \rightarrow I = \frac{V}{ZS + Zin}$$

Since antennas are often high-frequency devices, transmission line effects are often VERY important. That is, if the length L of the transmission line significantly alters Zin, then the current into the antenna from the source will be very small. Consequently, we will not be delivering power properly to the antenna. The same problems hold true in the receiving mode: a transmission line can skew impedance of the receiver sufficiently that almost no power is transferred from the antenna.

Hence, a thorough understanding of antenna theory requires an understanding of transmission lines. A great antenna can be hooked up to a great receiver, but if it is done with a length of transmission line at high frequencies, the system will not work properly.

Examples of common transmission lines include the coaxial cable, the microstrip line which commonly feeds patch/microstrip antennas, and the two wire line:

Coaxial Cable   Microstrip Line   Two-Wire Line

.
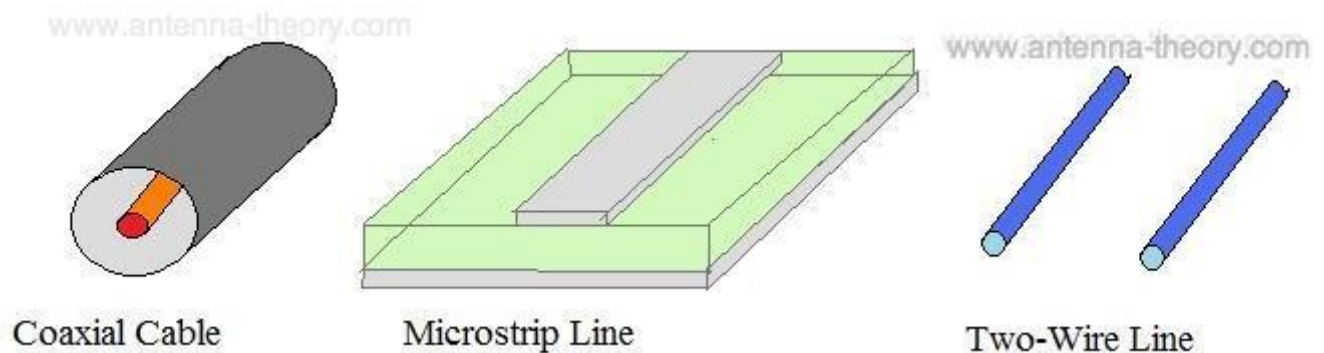
To understand transmission lines, we'll set up an equivalent circuit to model and analyze them. To start, we'll take the basic symbol for a transmission line of length L and divide it into small segments:

(a) Parallel-Wire Representation   (b) Divided into Small Sections of Length dz
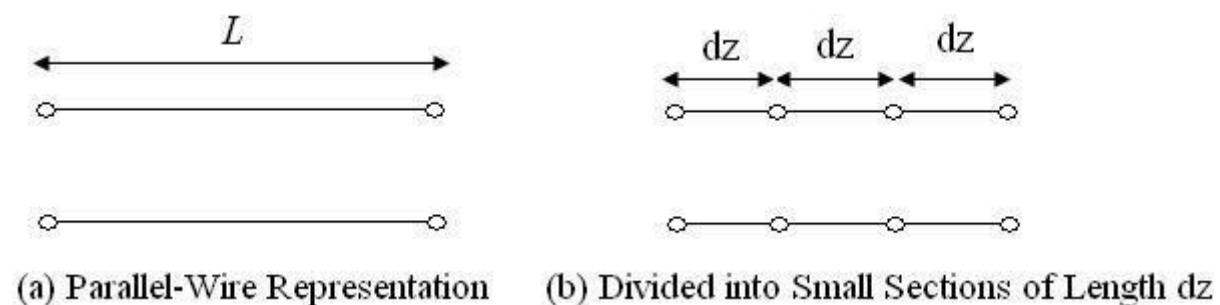
Then we'll model each small segment with a small series resistance, series inductance, shunt conductance, and shunt capcitance:
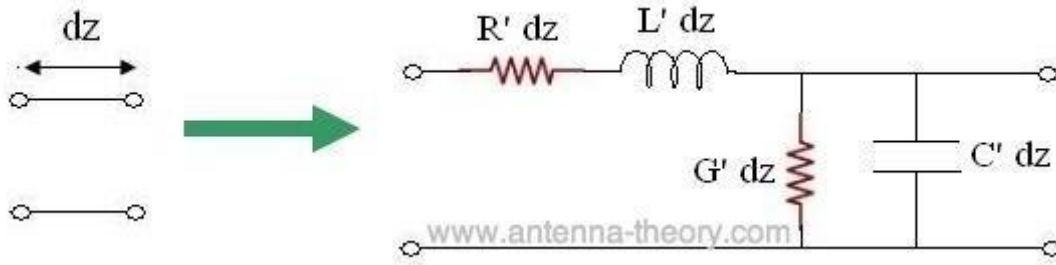
The parameters in the above figure are defined as follows:

R' - resistance per unit length for the transmission line (Ohms/meter)

L' - inductance per unit length for the tx line (Henries/meter)

G' - conductance per unit length for the tx line (Siemans/meter)

C' - capacitance per unit length for the tx line (Farads/meter)

We will use this model to understand the transmission line. All transmission lines will be represented via the above circuit diagram. For instance, the model for coaxial cables will differ from microstrip transmission lines only by their parameters R', L', G' and C'.

To get an idea of the parameters, R' would represent the d.c. resistance of one meter of the transmission line. The parameter G' represents the isolation between the two conductors of the transmission line. C' represents the capacitance between the two conductors that make up the tx line; L' represents the inductance for one meter of the tx line. These parameters can be derived for each transmission line. An example of deriving the paramters for a coaxial cable is given here.

Assuming the +z-axis is towards the right of the screen, we can establish a relationship between the voltage and current at the left and right sides of the terminals for our small section of transmission line:
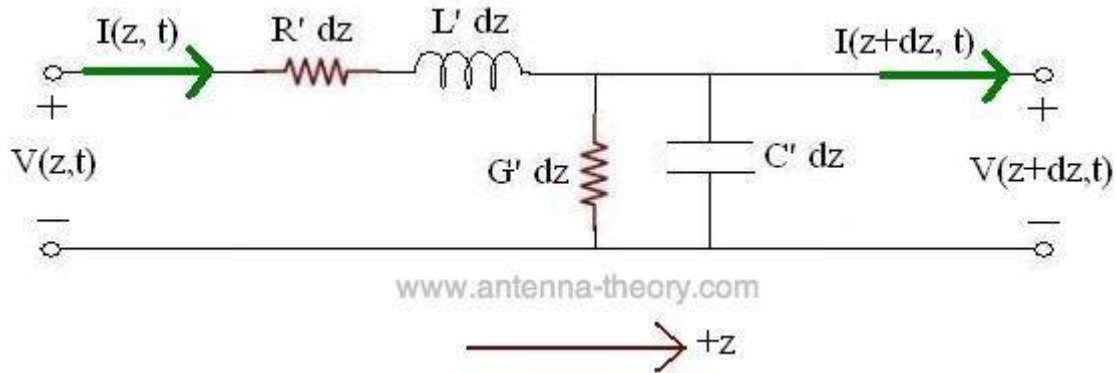
Using oridinary circuit theory, the relationship between the voltage and current on the left and right side of the transmission line segment can be derived:

$$\left[\frac{V(z+dz,t)-V(z,t)}{dz}\right] + R'I(z,t) + L'\frac{\partial I(z,t)}{\partial t} = 0$$

$$\left[\frac{I(z+dz,t)-I(z,t)}{dz}\right] + G'V(z,t) + C'\frac{\partial V(z,t)}{\partial t} = 0$$

Taking the limit as dz goes to zero, we end up with a set of differential equations that relates the voltage and current on an infinitesimal section of transmission line:

$$\frac{\partial V(z,t)}{\partial z} + R'I(z,t) + L'\frac{\partial I(z,t)}{\partial t} = 0$$

$$\frac{\partial I(z,t)}{\partial z} + G'V(z,t) + C'\frac{\partial V(z,t)}{\partial t} = 0$$

These equations are known as the telegraphers equations. Manipulation of these equations in phasor form allow for second order wave equations to be made for both V and I:

$$\frac{d^2V(z,t)}{dz^2} - \gamma^2 V(z,t) = 0$$

$$\frac{d^2I(z,t)}{dz^2} - \gamma^2 I(z,t) = 0$$

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

The solution of the above wave-equations will reveal the complex nature of transmission lines. Using ordinary differential equations theory, the solutions for the above differential equations are given by:

$$V(z,t) = V^+ e^{(j\omega t - \gamma z)} + V^- e^{(j\omega t + \gamma z)}$$

$$I(z,t) = I^+ e^{(j\omega t - \gamma z)} + I^- e^{(j\omega t + \gamma z)}$$

The solution is the sum of a forward traveling wave (in the +z direction) and a backward traveling wave (in the -z direction). In the above, V^+ is the amplitude of the forward traveling voltage wave, V^- is the amplitude of the backward traveling voltage wave, I^+ is the amplitude of the forward traveling current wave, and I^- is the amplitude of the backward traveling current wave.

4. THE INFINITESIMAL LINE

Consider the infinitesimal transmission line. It is recognized immediately that this line, in the limit may be considered as made up of cascaded infinitesimal T sections. The distribution of Voltage and Current are shown in hyperbolic form:

(4.1)

$$E = E_R \cosh\sqrt{Z_L Y_L}\, s + I_R\sqrt{\frac{Z_L}{Y_L}}\sinh\sqrt{Z_L Y_L}\, s$$

$$E = E_R \cosh\sqrt{Z_L Y_L}\, s + I_R\sqrt{\frac{Z_L}{Y_L}}\sinh\sqrt{Z_L Y_L}\, s$$

(4.2)

$$I = I_R \cosh\sqrt{Z_L Y_L}\, s + E_R\sqrt{\frac{Y_L}{Z_L}}\sinh\sqrt{Z_L Y_L}\, s$$

And shown in matrix form:

(4.3)

$$\begin{bmatrix} E \\ I \end{bmatrix} = \begin{bmatrix} \cosh\sqrt{Z_L Y_L}\,s & \sqrt{\frac{Z_L}{Y_L}}\sinh\sqrt{Z_L Y_L}\,s \\ \sqrt{\frac{Y_L}{Z_L}}\sinh\sqrt{Z_L Y_L}\,s & \cosh\sqrt{Z_L Y_L}\,s \end{bmatrix} \begin{bmatrix} E_R \\ I_R \end{bmatrix}$$

Where Z_L and Y_L are the series impedance and shunt admittance per unit length of line respectively.

Where the image impedance of the line is:

(4.4)

$$Z_{O_L} = \sqrt{\frac{Z_L}{Y_L}}$$

And the Propagation constant of the line is:

(4.5)

$$\gamma_L = \sqrt{Z_L Y_L}$$

And s is the distance to the point of observation, measured from the receiving end of the line.

Equations (4.1) and (4.2) are of the same form as equations (3.13) and (3.14) and are solutions to the wave equation.

Let us define a set of expressions such that:

(4.6)

$$Z_L = \frac{N}{S} Z_T \gamma_X$$

(4.7)

$$Y_L = \frac{N}{S} Y_T \gamma_Y$$

Where

$$\gamma_x = \frac{(\alpha-\beta)\gamma_T}{2} \quad \text{and} \quad \gamma_Y = \frac{2\,\gamma_T}{\alpha-\beta}\gamma_X = \frac{(\alpha-\beta)\gamma_T}{2} \quad \text{and} \quad \gamma_Y = \frac{2\,\gamma_T}{\alpha-\beta}$$

Also   note   that:

$$\gamma_T = \sqrt{\gamma_x\gamma_Y} \text{ which leads to } Z_L Y_L = \gamma_x\gamma_Y = \gamma_T^2$$

$\gamma_T = \sqrt{\gamma_X\gamma_Y}$ which leads to $Z_L Y_L = \gamma_X\gamma_Y = \gamma_T^2$

If we now substitute equations (4.6) and (4.7) into equations (4.4) and (4.5), and allowing $\qquad Z_r Y_r - 1/Z_T Y_T = 1 \qquad$ we have:

$$Y_L S = \gamma_r N\gamma_L S = \gamma_T N \qquad \text{And} \qquad Z_{O_L} = Z_r\frac{\sqrt{5}}{2}Z_{O_L} = Z_T\frac{\sqrt{5}}{2}$$

so by choosing $\qquad Z_r Y_r - 1/Z_T Y_T = 1 \qquad$ and then using equations (4.6) and (4.7) to find Z_L and $1/Z_L$ and Y_L, both Real, Imaginary or Complex, then equations (4.3) will be equivalent to equation (3.15) and equation (3.16).

So that the infinitesimal transmission line of distributed parameters, with Z and Y of the line as found from equations (4.6)and (4.7), a distance S from the generator, is now electrically equivalent to a line of N individual T sections whose $Z_r Y_r - 1/Z_T Y_T = 1$.

Quarter wave length

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

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

Matched load

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

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

for all l and all λ.

Short

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

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

Open

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

$$Z_{in}(l) = -jZ_0 \cot(\beta l)$$

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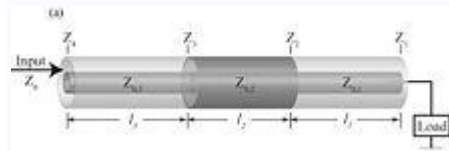
A simple example of stepped transmission line consisting of three segments.

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

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

where β_i is the wave number of the ith transmission line segment and l_i is the length of this segment, and Z_i is the front-end impedance that loads the ith segment.
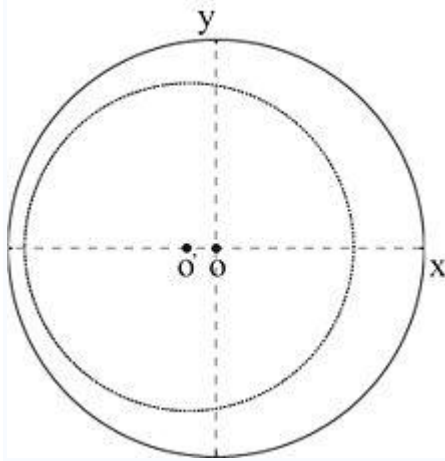
The impedance transformation circle along a transmission line whose characteristic impedance $Z_{0,i}$ is smaller than that of the input cable Z_0. And as a result, the impedance curve is off-centered towards the -x axis. Conversely, if $Z_{0,i} > Z_0$, the impedance curve should be off-centered towards the +x axis.

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

Practical types

Coaxial cable

Coaxial lines confine the electromagnetic wave to the area inside the cable, between the center conductor and the shield. The transmission of energy in the line occurs totally through the dielectric inside the cable between the conductors. Coaxial lines can therefore be bent and twisted (subject to limits) without negative effects, and they can be strapped to conductive supports without inducing unwanted currents in them. In radio-frequency applications up to a few gigahertz, the wave propagates in the transverse electric and magnetic mode (TEM) only, which means that the electric and magnetic fields are both perpendicular to the direction of propagation (the electric field is radial, and the magnetic field is circumferential). However, at frequencies for which the wavelength (in the dielectric) is significantly shorter than the circumference of the cable, transverse electric (TE) and transverse magnetic (TM) waveguide modes can also propagate.

When more than one mode can exist, bends and other irregularities in the cable geometry can cause power to be transferred from one mode to another.

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

Microstrip

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

Stripline

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

Balanced lines

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

Twisted pair

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

Single-wire line

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

Waveguide

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

Reflection coefficient

The reflection coefficient is used in physics and electrical engineering when wave propagation in a medium containing discontinuities is considered. A reflection coefficient describes either the amplitude or the intensity of a reflected wave relative to an incident wave. The reflection coefficient is closely related to the transmission coefficient.

Telecommunications

In telecommunications, the reflection coefficient is the ratio of the amplitude of the reflected wave to the amplitude of the incident wave. In particular, at a discontinuity in a transmission line, it is the complex ratio of the electric field strength of the reflected wave (E^-) to that of the incident wave (E^+). This is typically represented with a Γ (capital gamma) and can be written as:

$$\Gamma = \frac{E^-}{E^+}$$

The reflection coefficient may also be established using other field or circuit quantities.

The reflection coefficient can be given by the equations below, where Z_S is the impedance toward the source, Z_L is the impedance toward the load:
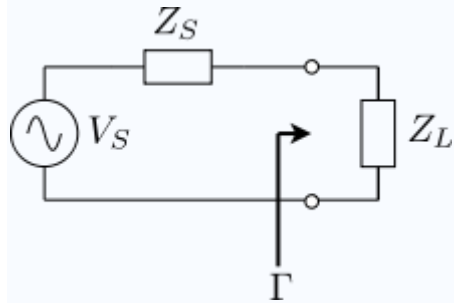
Simple circuit configuration showing measurement location of reflection coefficient.

$$\Gamma = \frac{Z_L - Z_S}{Z_L + Z_S}$$

Notice that a negative reflection coefficient means that the reflected wave receives a 180°, or π, phase shift.

The absolute magnitude (designated by vertical bars) of the reflection coefficient can be calculated from the standing wave ratio, SWR:

$$|\Gamma| = \frac{SWR - 1}{SWR + 1}$$

Insertion loss:

Insertion loss is a figure of merit for an electronic filter and this data is generally specified with a filter. Insertion loss is defined as a ratio of the signal level in a test configuration without the filter installed (V_1) to the signal level with the filter installed (V_2). This ratio is described in dB by the following equation:

$$\text{Insertion loss (dB)} = 10 \log_{10} \frac{V_1^2}{V_2^2} = 20 \log_{10} \frac{V_1}{V_2}$$

Filters are sensitive to source and load impedances so the exact performance of a filter in a circuit is difficult to precisely predict. Comparisons, however, of filter performance are possible if the insertion loss measurements are made with fixed

source and load impedances, and 50 Ω is the typical impedance to do this. This data is specified as common-mode or differential-mode. Common-mode is a measure of the filter performance on signals that originate between the power lines and chassis ground, whereas differential-mode is a measure of the filter performance on signals that originate between the two power lines.

Link with Scattering parameters

Insertion Loss (IL) is defined as follows:

$$IL = 10 \log_{10} \frac{|S_{21}|^2}{1 - |S_{11}|^2} \, \text{dB}$$

This definition results in a negative value for insertion loss, that is, it is really defining a gain, and a gain less than unity (i.e., a loss) will be negative when expressed in dBs. However, it is conventional to drop the minus sign so that an increasing loss is represented by an increasing positive number as would be intuitively expected

Recall the expressions for the voltage and current on the line (derived on the previous page):

$$V(z,t) = V^+ e^{\left(j\omega t - \gamma z\right)} + V^- e^{\left(j\omega t + \gamma z\right)}$$

$$I(z,t) = I^+ e^{\left(j\omega t - \gamma z\right)} + I^- e^{\left(j\omega t + \gamma z\right)}$$

If we plug this into equation [1] (note that z is fixed, because we are evaluating this at a specific point, the end of the transmission line), we obtain:

$$\frac{V^+ + V^-}{I^+ - I^-} = \frac{V^+ + V^-}{V^+ - V^-} Z_0 = Z_L$$

The ratio of the reflected voltage amplitude to that of the forward voltage amplitude is the voltage reflection coefficient. This can be solved for via the above equation:

$$\Gamma = \frac{V^-}{V^+} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

The reflection coefficient is usually denoted by the symbol gamma. Note that the magnitude of the reflection coefficient does not depend on the length of the line, only the load impedance and the impedance of the transmission line. Also, note that if ZL=Z0, then the line is "matched". In this case, there is no mismatch loss and all power is transferred to the load. At this point, you should begin to understand the importance of impedance matching: grossly mismatched impedances will lead to most of the power reflected away from the load.

Note that the reflection coefficient can be a real or a complex number.

Standing Waves

We'll now look at standing waves on the transmission line. Assuming the propagation constant is purely imaginary (lossless line), We can re-write the voltage and current waves as:

$$V(z) = V^+\left(e^{-j\beta z} + \Gamma e^{j\beta z}\right)$$

$$I(z) = \frac{V^+}{Z_0}\left(e^{-j\beta z} - \Gamma e^{j\beta z}\right)$$

If we plot the voltage along the transmission line, we observe a series of peaks and minimums, which repeat a full cycle every half-wavelength. If gamma equals 0.5 (purely real), then the magnitude of the voltage would appear as:

Similarly, if gamma equals zero (no mismatch loss) the magnitude of the voltage would appear as:

$|V(z)|$ $\Gamma = 0$

-1   -0.8   -0.6   -0.4   -0.2   0

← Wavelengths Away from Load   Load (z=0)

Finally, if gamma has a magnitude of 1 (this occurs, for instance, if the load is entirely reactive while the transmission line has a Z0 that is real), then the magnitude of the voltage would appear as:

One thing that becomes obvious is that the ratio of Vmax to Vmin becomes larger as the reflection coefficient increases. That is, if the ratio of Vmax to Vmin is one, then there are no standing waves, and the impedance of the line is perfectly matched to the load. If the ratio of Vmax to Vmin is infinite, then the magnitude of the reflection coefficient is 1, so that all power is reflected. Hence, this ratio, known as the Voltage Standing Wave Ratio (VSWR) or standing wave ratio is a measure of how well matched a transmission line is to a load. It is defined as:

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

Input impedance of a transmission line:
  Determine the input impedance of a transmission line of length L attached to a load (antenna) with impedance ZA. Consider the following circuit:

In low frequency circuit theory, the input impedance would simply be ZA. However, for high-frequency (or long) transmission lines, we know that the voltage and the current are given by:

$$V(z) = V^+\left(e^{-j\beta z} + \Gamma e^{j\beta z}\right)$$

$$I(z) = \frac{V^+}{Z_0}\left(e^{-j\beta z} - \Gamma e^{j\beta z}\right)$$

For simplicity, assume the transmission line is lossless, so that the propagation constant is purely imaginary. If we define z=0 to be at the terminals of the load or antenna, then we are interested in the ratio of the voltage to the current at location z=-L:

$$Z_{in}(z) = \frac{V(z)}{I(z)} = \frac{V^+\left(e^{-j\beta z} + \Gamma e^{j\beta z}\right)}{V^+\left(e^{-j\beta z} - \Gamma e^{j\beta z}\right)} Z_0$$

$$Z_{in}(z) = \left[\frac{e^{-j\beta z} + \Gamma e^{j\beta z}}{e^{-j\beta z} - \Gamma e^{j\beta z}}\right] Z_0$$

Using the definition for gamma (the voltage reflection coefficient), the above equation can be manipulated algebraically, and when evaluated at z=-L, we obtain:

$$Z_{in}(-L) = Z_0 \left[\frac{Z_A + jZ_0 \tan(\beta L)}{Z_0 + jZ_A \tan(\beta L)} \right]$$

This last equation is fundamnetal to understanding transmission lines. The input impedance of a load ZA is transformed by a transmission line as in the above equation. This equation can cause ZA to be transformed radically. An example will now be presented.

Example

Consider a voltage source, with generator impedance Zg, hooked to an antenna with impedance ZA via a transmission line. Suppose that Zg=50 Ohms, ZA=50 Ohms, Z0=200 Ohm, and that the line is a quarterwavelength long. How much power does the generator deliver?

Answer: The diagram for this problem is given in the following diagram:

(a) Actual Representation

(b) Equivalent Representation

The above diagram also shows the "equivalent circuit". The input impedance becomes:

$$\beta L = \left(\frac{2\pi}{\lambda}\right)\left(\frac{\lambda}{4}\right) = \frac{\pi}{2}$$

$$Z_{in}(-L) = Z_0\left[\frac{Z_A + jZ_0 \tan(\beta L)}{Z_0 + jZ_A \tan(\beta L)}\right]$$

$$= \left(200\left[\frac{50 + j(200)\tan\left(\frac{\pi}{2}\right)}{200 - j(50)\tan\left(\frac{\pi}{2}\right)}\right]\right) \rightarrow 200\frac{200}{50} = 800\Omega$$

Hence, the current that flows is given by:

$$I = \frac{V}{Z_g + Z_{in}} = \frac{V}{850} \ [Amps]$$

Note that if high frequency circuit theory was not taken into account, the current flow would have been V/100 Amps. This illustrates how transmission lines can upset the expected operation of high frequency circuits.

Quarter-Wave Transformer

Recall our formula for the input impedance of a transmission line of length L with characteristic impedance Z0 and connected to a load with impedance ZA:

$$Z_{in}(-L) = Z_0\left[\frac{Z_A + jZ_0 \tan(\beta L)}{Z_0 + jZ_A \tan(\beta L)}\right]$$

An interesting thing happens when the length of the line is a quarter of a wavelength:

$$Z_{in}\left(L=\frac{\lambda}{4}\right)=Z_0\left[\frac{Z_A+jZ_0\tan\left(\frac{2\pi}{\lambda}\frac{\lambda}{4}\right)}{Z_0+jZ_A\tan\left(\frac{2\pi}{\lambda}\frac{\lambda}{4}\right)}\right]$$

$$\boxed{Z_{in}\left(L=\frac{\lambda}{4}\right)=\frac{Z_0^2}{Z_A}}$$

The above equation is important: it states that by using a quarter-wavelength of transmission line, the impedance of the load (ZA) can be transformed via the above equation. The utility of this operation can be seen via an example.

Example. Match a load with impedance ZA=100 Ohms to be 50 Ohms using a quarter-wave transformer, as shown below.

Zin =50 Ohms   www.antenna-theory.com

Solution: The problem is to determine Z0 (the characteristic impedance of our quarter-wavelength transmission line) such that the 100 Ohm load is matched to 50 Ohms. By applying the above equation, the problem is simple:

$$Z_{in}\left(L=\frac{\lambda}{4}\right)=\frac{Z_0^2}{Z_A}$$

$$50=\frac{Z_0^2}{100}\Rightarrow Z_0=\sqrt{(50)(100)}=70.71\ \Omega$$

Hence, by using a transmission line with a characteristic impedance of 70.71 Ohms, the 100 Ohm load is matched to 50 Ohms. Hence, if a transmitter has an impedance of 50 Ohms and is trying to deliver power to the load (antenna), no power will be reflected back to the transmitter. In general, impedance matching is very important in RF/microwave circuit design. It is relatively simple at a single frequency, but becomes very difficult if wideband impedance matching is desired.

This technique is commonly employed with patch antennas. Circuits are printed as shown in the following figure. A 50 Ohm microstrip transmission line is matched to a patch antenna (impedance typically 200 Ohms or more) via a quarter-wavelength microstrip transmission line with the characteristic impedance chosen to match the load.

Because the quarter-wavelength transmission line is only a quarter-wavelength at a single frequency, this is a narrow-band matching technique. In the next section, we'll look at more uses of transmission lines.

Stub

In microwave and radio-frequency engineering, a stub is a length of transmission line or waveguide that is connected at one end only. The free end of the stub is either left open-circuit or (especially in the case of waveguides) short-circuited. Neglecting transmission line losses, the input impedance of the stub is purely reactive; either capacitive or inductive, depending on the electrical length of the

stub, and on whether it is open or short circuit. Stubs may thus be considered to be frequency-dependent capacitors and frequency-dependent inductors.

Because stubs take on reactive properties as a function of their electrical length, stubs are most common in UHF or microwave circuits where the line lengths are more manageable. Stubs are commonly used in antenna impedance matching circuits and frequency selective filters.

Smith charts can also be used to determine what length line to use to obtain a desired reactance.

Short circuited stub

The input impedance of a lossless short circuited line is,

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

where j is the imaginary unit, Z_0 is the characteristic impedance of the line, β is the phase constant of the line, and l is the physical length of the line.

Thus, depending on whether tan(βl) is positive or negative, the stub will be inductive or capacitive, respectively.

The Length of a stub to act as a capacitor C at an angular frequency of ω is then given by:

$$l = \frac{1}{\beta}\left[(n+1)\pi - \arctan\left(\frac{1}{\omega C Z_0}\right)\right]$$

The length of a stub to act as an inductor L at the same frequency is given by:

$$l = \frac{1}{\beta}\left[n\pi + \arctan\left(\frac{\omega L}{Z_0}\right)\right]$$

Open circuited stub

The input impedance of a lossless open circuit stub is given by

$$Z_{OC} = -jZ_0 \cot(\beta l)$$

It follows that whether cot(βl) is positive or negative, the stub will be capacitive or inductive, respectively.

The length of an open circuit stub to act as an Inductor L at an angular frequency of ω is:

$$l = \frac{1}{\beta} \left[(n+1)\pi - \text{arccot} \left(\frac{\omega L}{Z_0} \right) \right]$$

The length of an open circuit stub to act as a capacitor C at the same frequency is:

$$l = \frac{1}{\beta} \left[n\pi + \text{arccot} \left(\frac{1}{\omega C Z_0} \right) \right]$$

Stub matching

In a stripline circuit, a stub may be placed just before an output connector to compensate for small mismatches due to the device's output load or the connector itself.

Stubs can be used to match a load impedance to the transmission line characteristic impedance. The stub is positioned a distance from the load. This distance is chosen so that at that point the resistive part of the load impedance is made equal to the resistive part of the characteristic impedance by impedance transformer action of the length of the main line. The length of the stub is chosen so that it exactly cancels the reactive part of the presented impedance. That is, the stub is made capacitive or inductive according to whether the main line is presenting an inductive or capacitive impedance respectively. This is not the same as the actual impedance of the load since the reactive part of the load impedance will be subject to impedance transformer action as well as the resistive part. Matching stubs can be made adjustable so that matching can be corrected on test.

single stub will only achieve a perfect match at one specific frequency. For wideband matching several stubs may be used spaced along the main transmission line. The resulting structure is filter-like and filter design techniques are applied. For instance, the matching network may be designed as a Chebyshev filter but is optimised for impedance matching instead of passband transmission. The resulting transmission function of the network has a passband ripple like the Chebyshev filter, but the ripples never reach 0dB insertion loss at any point in the passband, as they would do for the standard filter.

Single stub impedance matching:

- The load should be matched to the characteristic impedance of the line so that as much power as possible is transmitted from the generator to the load for radio-frequency power transmission.

- The lines should be matched because reflections from mismatched loads and junctions will result in echoes and will distort the information-carrying signal for information transmission.

- Short-circuited (instead of open-circuited) stubs are used for impedance-matching on transmission lines.

- Single-stub method for impedance matching : an arbitrary load impedance can be matched to a transmission line by placing a single short-circuited stub in parallel with the line at a suitable location (Fig.8-12).

FIGURE 8-12 Impedance matching by single-stub method.

$Y_0 = 1/R_0.$

- : the input admittance at BB' looking toward the load without the stub.

- Our impedance- (or admittance-) matching problem : to determine the location d and the length of the stub such that

$$Y_i = Y_0 = Y_B + Y_s, \tag{8-114}$$

  Where $Y_0 = 1/R_0$

- In terms of normalized admittance, the above equation becomes

$$1 = y_B + y_s, \tag{8-115}$$

- Y_s purely imaginary (the input admittance of a short-circuited stub is purely susceptive).

- Thus, the above equation can be satisfied only if

$$y_B = 1 + jb_B \qquad\qquad\qquad (8\text{-}116)$$

and

$$y_s = -jb_B, \qquad\qquad\qquad (8\text{-}117)$$

where b_B can be either positive or negative.

- Our objectives : (1) to find the length d such that the admittance Y_B of the load section looking to the right of terminal BB' has a unitary real part, and (2) to find the length of the stub required to cancel the imaginary part.

Smith Chart:
The Smith Chart is a fantastic tool for visualizing the impedance of a transmission line and antenna system as a function of frequency. Smith Charts can be used to increase understanding of transmission lines and how they behave from an impedance viewpoint. Smith Charts are also extremely helpful for impedance matching, as we will see. The Smith Chart is used to display a real antenna's impedance when measured on a Vector Network Analyzer (VNA).

Smith Charts were originally developed around 1940 by Phillip Smith as a useful tool for making the equations involved in transmission lines easier to manipulate. See, for instance, the input impedance equation for a load attached to a transmission line of length L and characteristic impedance Z0. With modern computers, the Smith Chart is no longer used to the simplify the calculation of transmission line equatons; however, their value in visualizing the impedance of an antenna or a transmission line has not decreased.

The Smith Chart is shown in Figure 1. A larger version is shown here.

Figure 1. The basic Smith Chart.

Figure 1 should look a little intimidating, as it appears to be lines going everywhere. There is nothing to fear though. We will build up the Smith Chart from scratch, so that you can understand exactly what all of the lines mean. In fact, we are going to learn an even more complicated version of the Smith Chart known as the immitance Smith Chart, which is twice as complicated, but also twice as useful. But for now, just admire the Smith Chart and its curvy elegance.

This section of the antenna theory site will present an intro to the Smith Chart basics.

Smith Chart Tutorial

We'll now begin to explain the Smith Chart. The Smith Chart displays the complex reflection coefficient, in polar form, for an arbitrary impedance (we'll call the impedance ZL or the load impedance). For a primer on complex math, click here.

Recall that the complex reflection coefficient (Γ) for an impedance ZL attached to a transmission line with characteristic impedance Z0 is given by:

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}$$ [1]

For this tutorial, we will assume Z0 is 50 Ohms, which is often, but not always the case.

The complex reflection coefficient, or Γ, must have a magnitude between 0 and 1. As such, the set of all possible values for Γ must lie within the unit circle:

Figure 2. The Complex Reflection Coefficient must lie somewhere within the unit circle.

In Figure 2, we are plotting the set of all values for the complex reflection coefficient, along the real and imaginary axis. The center of the Smith Chart is the

point where the reflection coefficient is zero. That is, this is the only point on the smith chart where no power is reflected by the load impedance.

The outer ring of the Smith Chart is where the magnitude of Γ is equal to 1. This is the black circle in Figure 1. Along this curve, all of the power is reflected by the load impedance.

Normalized Load Impedance

To make the Smith Chart more general and independent of the characteristic impedance Z0 of the transmission line, we will normalize the load impedance ZL by Z0 for all future plots:

$$z_L = \frac{Z_L}{Z_0} = \frac{1+\Gamma}{1-\Gamma}$$
[1]

Equation [1] doesn't affect the reflection coefficient (Γ). It is just a convention that is used everywhere.

Constant Resistance Circles

for a given normalized load impedance zL, we can determine Γ and plot it on the Smith Chart.

Now, suppose we have the normalized load impedance given by:

$$z_1 = 1 + iY$$
[2]

In equation [2], Y is any real number. What would the curve corresponding to equation [2] look like if we plotted it on the Smith Chart for all values of Y? That is, if we plotted z1 = 1 + 0*i, and z1 = 1 + 10*i, z1 = 1 - 5*i, z1 = 1 - .333*i, ... and any possible value for Y that you could think of, what is the resulting curve? The answer is shown in Figure 1:

Figure 1. Constant Resistance Circle for zL=1 on Smith Chart.

In Figure 1, the outer blue ring represents the boundary of the smith chart. The black curve is a constant resistance circle: this is where all values of $z1 = 1 + i*Y$ will lie on. Several points are plotted along this curve, $z1 = 1$, $z1 = 1 + i*2$, and $zL = 1 - i*4$.

Suppose we want to know what the curve $z2 = 0.3 + i*Y$ looks like on the Smith Chart. The result is shown in Figure 2:

Figure 2. Constant Resistance Circle for zL=0.3 on Smith Chart.

In Figure 2, the black ring represents the set of all impedances where the real part of z2 equals 0.3. A few points along the circle are plotted. We've left the resistance circle of 1.0 in red on the Smith Chart.

These circles are called constant resistance curves. The real part of the load impedance is constant along each of these curves. We'll now add several values for the constant resistance, as shown in Figure 3:

Figure 3. Constant Resistance Circles on Smith Chart.

In Figure 3, the zL=0.1 resistance circle has been added in purple. The zL=6 resistance circle has been added in green, and zL=2 resistance circle is in black.

look at the set of curves defined by zL = R + iY, where Y is held constant and R varies from 0 to infinity. Since R cannot be negative for antennas or passive devices, we will restrict R to be greater than or equal to zero. As a first example, let zL = R + i. The curve defined by this set of impedances is shown in Figure 1:

Figure 1. Constant Reactance Curve for zL = R + i*1.

The resulting curve zL = R + i is plotted in green in Figure 1. A few points along the curve are illustrated as well. Observe that zL = 0.3 + i is at the intersection of the Re[zL] = 0.3 circle and the Im[zL]=1 curve. Similarly, observe that the zL = 2 + i point is at the intersection of the Re[zL]=2 circle and the Im[zL]=1 curve. (For a quick reminder of real and imaginary parts of complex numbers, see complex math primer.)

The constant reactance curve, defined by Im[zL]=-1 is shown in Figure 2:

Figure 2. Constant Reactance Curve for zL = R - i.

The resulting curve for Im[zL]=-1 is plotted in green in Figure 2. The point zL=1-i is placed on the Smith Chart, which is at the intersection of the Re[zL]=1 circle and the Im[zL]=-1 curve.

An important curve is given by Im[zL]=0. That is, the set of all impedances given by zL = R, where the imaginary part is zero and the real part (the resistance) is greater than or equal to zero. The result is shown in Figure 3:

Figure 3. Constant Reactance Curve for zL=R.

The reactance curve given by Im[zL]=0 is a straight line across the Smith Chart. There are 3 special points along this curve. On the far left, where zL = 0 + i0, this is the point where the load is a short circuit, and thus the magnitude of Γ is 1, so all power is reflected.

In the center of the Smith Chart, we have the point given by zL = 1. At this location, Γ is 0, so the load is exactly matched to the transmission line. No power is reflected at this point.

The point on the far right in Figure 3 is given by zL = infinity. This is the open circuit location. Again, the magnitude of Γ is 1, so all power is reflected at this point, as expected.

Finally, we'll add a bunch of constant reactance curves on the Smith Chart, as shown in Figure 4.

Figure 4. Smith Chart with Reactance Curves and Resistance Circles.

In Figure 4, we added constant reactance curves for Im[zL]=2, Im[zL]=5, Im[zL]=0.2, Im[zL]=0.5, Im[zL]=-2, Im[zL]=-5, Im[zL]=-0.2, and Im[zL] = -0.5.

Figure 4 shows the fundamental curves of the Smith Chart.

Applications of smith Chart:

Plotting an impedance

Measurement of VSWR

Measurement of reflection coefficient (magnitude and phase)

Measurement of input impedance of the line

The waves guided or directed by the guided structures are called guided wave. In general wave equations are derived from Maxwell's equation. To obtain the solution of this problem it is essential to apply certain restrictions or boundary conditions to the Maxwell's equation.

Maxwell's Equations are a set of four vector-differential equations that govern all of electromagnetics (except at the quantum level, in which case we as antenna people don't care so much). They were first presented in a complete form by James Clerk Maxwell back in the 1800s. He didn't come up with them all on his own, but did add the displacement current term to Ampere's law which made them complete.

The four equations (written only in terms of E and H, the electric field and the magnetic field), are given below.

$$\nabla \cdot \mathbf{E} = \frac{\rho_v}{\varepsilon} \qquad \text{(Gauss' Law)}$$

$$\nabla \cdot \mathbf{H} = 0 \qquad \text{(Gauss' Law for Magnetism)}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \qquad \text{(Faraday's Law)}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Ampere's Law)}$$
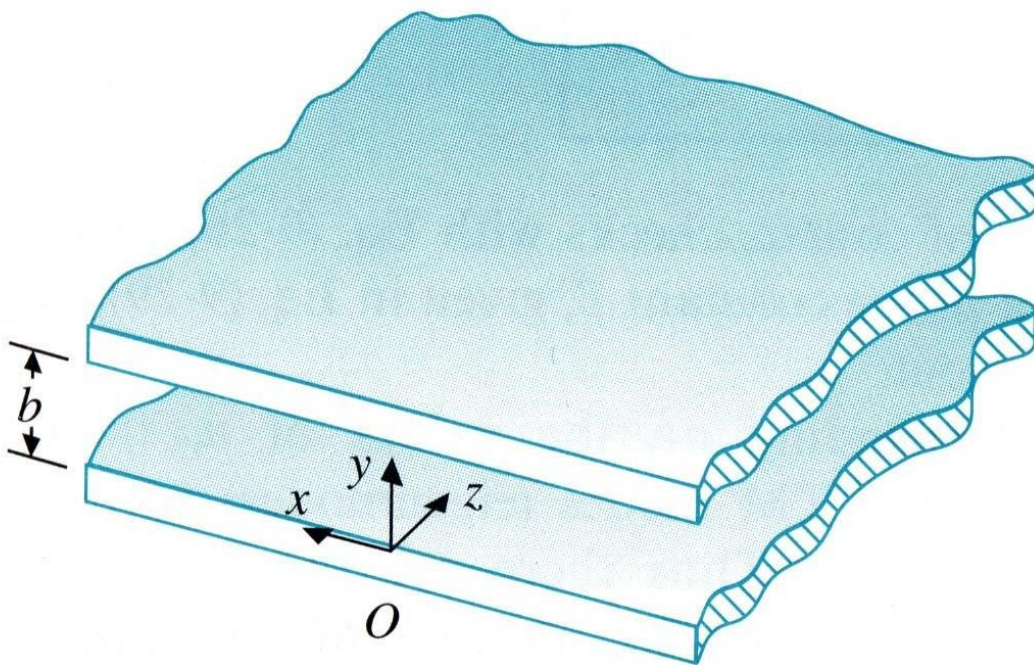
In Gauss' law, ρ_v is the volume electric charge density, J is the electric current density (in Amps/meter-squared), ε is the permittivity and μ is the permeability.

The good news about this is that all of electromagnetics is summed up in these 4 equations. The bad news is that no matter how good at math you are, these can only be solved with an analytical solution in extremely simple cases. Antennas don't present a very simple case, so these equations aren't used a whole lot in antenna theory (except for numerical methods, which numerically solve these approximately using a whole lot of computer power).

The last two equations (Faraday's law and Ampere's law) are responsible for electromagnetic radiation. The curl operator represents the spatial variation of the fields, which are coupled to the time variation. When the E-field travels, it is

altered in space, which gives rise to a time-varying magnetic field. A time-varying magnetic field then varies as a function of location (space), which gives rise to a time varying electric field. These equations wrap around each other in a sense, and give rise to a wave equation. These equations predict electromagnetic radiation as we understand it.

FIGURE 9-3   An infinite parallel-plate waveguide.

Consider a parallel-plate waveguide of two perfectly conducting plates separated by a distance b and filled with a dielectric medium having constitutive parameter as shown in Fig. 9-3. The plates are assumed to be infinite in extent in the x-direction. (Fields do not vary in the x-direction.)
   a) Obtain the time-harmonic field expressions for TM modes in the guide.
   b) Determine the cutoff frequency.
   a) For TM modes,
   - Eq. becomes

$$\frac{d^2 E_z^0(y)}{dy^2} + h^2 E_z^0(y) = 0. \tag{9-44}$$

The general solution for Eq. (9-44) :
  - Boundary conditions (The tangential component of the electric field must vanish on the surface of the perfectly conducting plates.) :
    (i) At y=0 Ez=0
    (ii) At y=b Ez=0
  - The value of the eigenvalue h :

$$h = \frac{n\pi}{b}, \qquad n = 1, 2, 3, \dots . \tag{9-46}$$

Types of propagation:
  TE waves
  TM waves
  TEM waves

Transverse electric(TE) wave has the magnetic field component in the direction of propagation, but no component of the electric field in the same direction. Hence the TE waves also known as M –waves or H-waves.

Transverse magnetic(TM) wave has the electric field in the direction of propagation, but no component of the magnetic field in the same direction. Hence the TM waves are also called E-waves.

Transverse electro magnetic(TEM) wave:
  No field in the direction of propagation

Attenuation of parallel plane guides:
  When the electromagnetic wave propagates through the wave guide, the amplitude of the fields or the signal strength of the wave decreases as the distance from the source increases. This is because when the wave strike the walls of the guide, the loss in the power takes place The attenuation factor is denoted by α.

A=Power lost per unit length/2*power transmitted

Attenuation due to finite wall conductivity is inversely proportional to the square root of wall conductivity, but depends on the mode and the frequency in a complicated way.

Attenuation due to wall losses in rectangular copper waveguide : Figure 9-7.

 TE10 mode has the lowest attenuation in a rectangular waveguide.

The attenuation constant increases rapidly toward infinity as the operating frequency approaches the cutoff frequency.

Causes for attenuation in waveguides : lossy dielectric and imperfectly conducting walls.

FIGURE 9-7 Attenuation due to wall losses in rectangular copper waveguide for TE_{10} and TM_{11} modes. $a = 2.29$ (cm), $b = 1.02$ (cm).

Cut-off frequency:

The frequency at which wave motion ceases is called cut-off frequency

$$f_c = \frac{n}{2b\sqrt{\mu\epsilon}} \quad \text{(Hz)}, \qquad (9\text{-}51)$$

Propagation Constant:

$$\gamma = \sqrt{\left(\frac{n\pi}{b}\right)^2 - \omega^2 \mu \epsilon}. \qquad \text{(9-50)}$$

Wave impedance:

It is the ratio of the component of the electric field to that of magnetic field.
Wave impedance for TEM wave

$$Z_{\text{TEM}} = \frac{E_x^0}{H_y^0} = \frac{j\omega\mu}{\gamma_{\text{TEM}}} = \frac{\gamma_{\text{TEM}}}{j\omega\epsilon}, \qquad \text{(9-19)}$$

Wave impedance for TM and TE wave

$$Z_{\text{TM}} = \frac{E_x^0}{H_y^0} = -\frac{E_y^0}{H_x^0} = \frac{\gamma}{j\omega\epsilon} \quad (\Omega). \qquad \text{(9-23)}$$

$$Z_{\text{TE}} = \frac{\eta}{\sqrt{1 - (f_c/f)^2}} \quad (\Omega), \qquad \text{(9-39)}$$

Phase velocity:

It is the velocity at which energy propagates along a wave guide

$$u_p = \frac{\omega}{\beta} \quad (\text{m/s}). \qquad \text{(7-58)}$$

- The phase velocity and the wave impedance for TEM waves are independent of the frequency of the waves.
- TEM waves cannot exist in a single-conductor hollow (or dielectric-filled) waveguide of any shape.

Rectangular waveguides are the one of the earliest type of the transmission lines. They are used in many applications. A lot of components such as isolators, detectors, attenuators, couplers and slotted lines are available for various standard waveguide bands between 1 GHz to above 220 GHz.

A rectangular waveguide supports TM and TE modes but not TEM waves because we cannot define a unique voltage since there is only one conductor in a rectangular waveguide. The shape of a rectangular waveguide is as shown below. A material with permittivity e and permeability m fills the inside of the conductor.

A rectangular waveguide cannot propagate below some certain frequency. This frequency is called the cut-off frequency.

Here, we will discuss TM mode rectangular waveguides and TE mode rectangular waveguides separately. Let's start with the TM mode.

TM Modes

Consider the shape of the rectangular waveguide above with dimensions a and b (assume a>b) and the parameters e and m. For TM waves $H_z = 0$ and E_z should be solved from equation for TM mode;

$$\tilde{N}^2_{xy}\, E_z^{\,0} + h^2\, E_z^{\,0} = 0$$

Since $E_z(x,y,z) = E_z^0(x,y)e^{-gz}$, we get the following equation,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + h^2\right) E_z^0(x,y) = 0$$

If we use the method of separation of variables, that is $E_z^0(x,y) = X(x).Y(y)$ we get,

$$-\frac{1}{X(x)}\frac{d^2 X(x)}{dx^2} = \frac{1}{Y(y)}\frac{d^2 Y(y)}{dy^2} + h^2$$

Since the right side contains x terms only and the left side contains y terms only, they are both equal to a constant. Calling that constant as k_x^2, we get;

$$\frac{d^2 X(x)}{dx^2} + k_x^2 X(x) = 0$$

$$\frac{d^2 Y(y)}{dy^2} + k_y^2 Y(y) = 0$$

where $k_y^2 = h^2 - k_x^2$

Now, we should solve for X and Y from the preceding equations. Also we have the boundary conditions of;

$E_z^0(0,y)=0$

$E_z^0(a,y)=0$

$E_z^0(x,0)=0$

$E_z^0(x,b)=0$

From all these, we conclude that

X(x) is in the form of sin $k_x x$, where k_x=mp/a, m=1,2,3,…

Y(y) is in the form of sin $k_y y$, where k_y=np/b, n=1,2,3,…

So the solution for $E_z^0(x,y)$ is

$$E_z^0(x,y) = E_0 \sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$ (V/m)

From $k_y^2 = h^2 - k_x^2$, we have;

$$\boxed{h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

For TM waves, we have

$$H_x^0 = \frac{jw\varepsilon}{h^2}\frac{\partial E_z^0}{\partial y}$$

$$H_y^0 = -\frac{jw\varepsilon}{h^2}\frac{\partial E_z^0}{\partial x}$$

$$E_x^0 = -\frac{\gamma}{h^2}\frac{\partial E_z^0}{\partial x}$$

$$E_y^0 = -\frac{\gamma}{h^2}\frac{\partial E_z^0}{\partial y}$$

From these equations, we get

$$E_x^0(x,y) = -\frac{\gamma}{h^2}\left(\frac{m\pi}{a}\right)E_0\cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

$$E_y^0(x,y) = -\frac{\gamma}{h^2}\left(\frac{n\pi}{b}\right)E_0\sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)$$

$$H_x^0(x,y) = \frac{jw\varepsilon}{h^2}\left(\frac{n\pi}{b}\right)E_0\sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)$$

$$H_y^0(x,y) = -\frac{jw\varepsilon}{h^2}\left(\frac{m\pi}{a}\right)E_0\cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

where

$$\gamma = j\beta = j\sqrt{w^2\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}$$

Here, m and n represent possible modes and it is designated as the TM$_{mn}$ mode. m denotes the number of half cycle variations of the fields in the x-direction and n denotes the number of half cycle variations of the fields in the y-direction.

When we observe the above equations we see that for TM modes in rectangular waveguides, neither m nor n can be zero. This is because of the fact that the field expressions are identically zero if either m or n is zero. Therefore, the lowest mode for rectangular waveguide TM mode is TM_{11} .

Here, the cut-off wave number is

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

and therefore,

$$\beta = \sqrt{k^2 - k_c^2}$$

The cut-off frequency is at the point where g vanishes. Therefore,

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \; (Hz)$$

Since l=u/f, we have the cut-off wavelength,

$$\lambda_c = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}} \; (m)$$

At a given operating frequency f, only those frequencies, which have $f_c<f$ will propagate. The modes with $f<f_c$ will lead to an imaginary b which means that the field components will decay exponentially and will not propagate. Such modes are called cut-off or evanescent modes.

The mode with the lowest cut-off frequency is called the dominant mode. Since TM modes for rectangular waveguides start from TM_{11} mode, the dominant frequency is

$$(f_c)_{11} = \frac{1}{2\sqrt{\varepsilon\mu}} \sqrt{\left(\frac{1}{a}\right)^2 + \left(\frac{1}{b}\right)^2} \ (Hz)$$

The wave impedance is defined as the ratio of the transverse electric and magnetic fields. Therefore, we get from the expressions for E_x and H_y (see the equations above);

$$Z_{TM} = \frac{E_x}{H_y} = \frac{\gamma}{jw\varepsilon} = \frac{j\beta}{jw\varepsilon} \Rightarrow Z_{TM} = \frac{\beta\eta}{k}$$

The guide wavelength is defined as the distance between two equal phase planes along the waveguide and it is equal to

$$\boxed{\lambda_g = \frac{2\pi}{\beta}} > \frac{2\pi}{k} = \lambda$$

which is thus greater than l, the wavelength of a plane wave in the filling medium.

The phase velocity is

$$u_p = \frac{w}{\beta} > \frac{w}{k} = \frac{1}{\sqrt{\mu\varepsilon}}$$

which is greater than the speed of light (plane wave) in the filling material.

Attenuation for propagating modes results when there are losses in the dielectric and in the imperfectly conducting guide walls. The attenuation constant due to the losses in the dielectric can be found as follows:

$$\gamma = j\beta = j\sqrt{k^2 - k_c^2} = jk\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu}\sqrt{\varepsilon + \frac{\sigma}{jw}}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

TE Modes

Consider again the rectangular waveguide below with dimensions a and b (assume a>b) and the parameters e and m.

For TE waves $E_z = 0$ and H_z should be solved from equation for TE mode;

$$\tilde{N}^2_{xy} H_z + h^2 H_z = 0$$

Since $H_z(x,y,z) = H^0_z(x,y)e^{-gz}$, we get the following equation,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + h^2\right)H^0_z(x,y) = 0$$

If we use the method of separation of variables, that is $H_z^0(x,y) = X(x).Y(y)$ we get,

$$-\frac{1}{X(x)}\frac{d^2 X(x)}{dx^2} = \frac{1}{Y(y)}\frac{d^2 Y(y)}{dy^2} + h^2$$

Since the right side contains x terms only and the left side contains y terms only, they are both equal to a constant. Calling that constant as k_x^2, we get;

$$\frac{d^2 X(x)}{dx^2} + k_x^2 X(x) = 0$$

$$\frac{d^2 Y(y)}{dy^2} + k_y^2 Y(y) = 0$$

where $k_y^2 = h^2 - k_x^2$

Here, we must solve for X and Y from the preceding equations. Also we have the following boundary conditions:

$$\frac{\partial H^0_z}{\partial x} = 0 (E_y = 0) \quad \text{at x=0}$$

$$\frac{\partial H^0_z}{\partial x} = 0 (E_y = 0) \quad \text{at x=a}$$

$$\frac{\partial H^0_z}{\partial y} = 0 (E_x = 0) \quad \text{at y=0}$$

$$\frac{\partial H_z^0}{\partial y} = 0 (E_x = 0)$$

at y=b

From all these, we get

$$H_z^0(x,y) = H_0 \cos\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right) \quad \text{(A/m)}$$

From $k_y^2 = h^2 - k_x^2$, we have;

$$\boxed{h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

For TE waves, we have

$$H_x^0 = -\frac{\gamma}{h^2}\frac{\partial H_z^0}{\partial x}$$

$$H_y^0 = -\frac{\gamma}{h^2}\frac{\partial H_z^0}{\partial y}$$

$$E_x^0 = -\frac{j w \mu}{h^2}\frac{\partial H_z^0}{\partial y}$$

$$E_y^0 = -\frac{j w \mu}{h^2}\frac{\partial H_z^0}{\partial x}$$

From these equations, we obtain

$$E_x^0(x,y) = \frac{j w \mu}{h^2}\left(\frac{n\pi}{b}\right)H_0 \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

$$\boxed{E_y^0(x,y) = -\frac{j w \mu}{h^2}\left(\frac{m\pi}{a}\right)H_0 \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)}$$

$$H_x^0(x,y) = \frac{\gamma}{h^2}\left(\frac{m\pi}{a}\right)H_0 \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)$$

$$H_y^0(x,y) = \frac{\gamma}{h^2}\left(\frac{n\pi}{b}\right)H_0 \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

where

$$\gamma = j\beta = j\sqrt{w^2\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}$$

As explained before, m and n represent possible modes and it is shown as the TE$_{mn}$ mode. m denotes the number of half cycle variations of the fields in the x-direction and n denotes the number of half cycle variations of the fields in the y-direction.

Here, the cut-off wave number is

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

and therefore,

$$\beta = \sqrt{k^2 - k_c^2}$$

The cut-off frequency is at the point where g vanishes. Therefore,

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}\ (Hz)$$

Since l=u/f, we have the cut-off wavelength,

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

At a given operating frequency f, only those frequencies, which have $f > f_c$ will propagate. The modes with $f < f_c$ will not propagate.

The mode with the lowest cut-off frequency is called the dominant mode. Since TE_{10} mode is the minimum possible mode that gives nonzero field expressions for rectangular waveguides, it is the dominant mode of a rectangular waveguide with $a > b$ and so the dominant frequency is

$$(f_c)_{10} = \frac{1}{2a\sqrt{\mu\varepsilon}}\ (Hz)$$

The wave impedance is defined as the ratio of the transverse electric and magnetic fields. Therefore, we get from the expressions for E_x and H_y (see the equations above);

$$Z_{TE} = \frac{E_x}{H_y} = \frac{jw\mu}{\gamma} = \frac{jw\mu}{j\beta} \Rightarrow Z_{TE} = \frac{k\eta}{\beta}$$

The guide wavelength is defined as the distance between two equal phase planes along the waveguide and it is equal to

$$\boxed{\lambda_g = \frac{2\pi}{\beta}} > \frac{2\pi}{k} = \lambda$$

which is thus greater than l, the wavelength of a plane wave in the filling medium.

The phase velocity is

$$u_p = \frac{w}{\beta} > \frac{w}{k} = \frac{1}{\sqrt{\mu\varepsilon}}$$

which is greater than the speed of the plane wave in the filling material.

The attenuation constant due to the losses in the dielectric is obtained as follows:

$$\gamma = j\beta = j\sqrt{k^2 - k_c^2} = jk\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu}\sqrt{\varepsilon + \frac{\sigma}{jw}}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

After some manipulation, we get

$$\alpha_d = \frac{\sigma\eta}{2\sqrt{1-\left(\dfrac{f_c}{f}\right)^2}} = \frac{k^2 \tan\delta}{2\beta}$$

Example:

Consier a length of air-filled copper X-band waveguide, with dimensions a=2.286cm, b=1.016cm. Find the cut-off frequencies of the first four propagating modes.

Solution:

From the formula for the cut-off frequency

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \xrightarrow{\text{air-filled}} \frac{c}{2}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \ (Hz)$$

FIGURE 9-5 Field lines for TE_{10} mode in rectangular waveguide.

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, fc. 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:

$$f_c = \frac{1}{2a\sqrt{\mu\varepsilon}} = \frac{c}{2a} \qquad \text{[Rectangular Cross Section]}$$

In the above, c is the speed of light within the waveguide, mu is the permeability of the material that fills the waveguide, and epsilon 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:

$$f_c = \frac{1.8412}{2\pi a\sqrt{\mu\varepsilon}} = \frac{1.8412c}{2\pi a} \qquad \text{[Circular Cross Section]}$$

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 Ez=0) or TM ('transverse magnetic' - which indicates that the H-field is orthogonal to the axis of the waveguide, so Hz = 0). The modes are further classified as TEij, 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 Ez and Hz are zero) mode. Their 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, 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 D and the corresponding electric field E as:

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

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

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 Hz and Ez are equal to zero. Hence, what field configurations that are allowed will be classified as either TM (Transverse Magnetic, in which Hz=0) and TE (Transverse Electric, in which Ez=0). The reason that waves cannot be TEM (Transverse Electromagnetic, Hz=Ez=0) will be shown towards the end of this derivation.

To perform our analysis, we'll assume that Ez=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:

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

$$E_y = \frac{1}{\varepsilon} \frac{\partial F_z}{\partial x}$$

$$E_z = 0$$

$$H_x = -j \frac{1}{\omega\mu\varepsilon} \frac{\partial^2 F_z}{\partial x \partial z}$$

$$H_y = -j \frac{1}{\omega\mu\varepsilon} \frac{\partial^2 F_z}{\partial y \partial z}$$

$$H_z = -j \frac{1}{\omega\mu\varepsilon} \left(\frac{\partial^2}{\partial z^2} + k^2 \right) F_z$$

Therefore, if we can find Fz (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:

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

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):

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

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

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

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

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

$$F_z(x, y, z) = X(x)Y(y)Z(z)$$ [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 Fz (equation [3]) into equation [2], and we end up with:

$$\frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} = -k^2$$ [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):

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

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

$$X'' + k_x^2 X = 0$$
$$Y'' + k_y^2 Y = 0$$
$$Z'' + k_z^2 Z = 0$$ [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:

$$X(x) = c_1 \cos(k_x x) + c_2 \sin(k_x x)$$

$$Y(y) = c_3 \cos(k_y y) + c_4 \sin(k_y y)$$

$$Z(z) = c_5 e^{jk_z z} + c_6 e^{-jk_z z}$$

[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 c5=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 Fz can be written as:

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

[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 Electric fields at any perfect conductor must be zero (why? because $J = \sigma E$, 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 Ex 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, Ey must be zero when x=0 and when x=a (independent of x and z). We will calculate Ex:

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

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

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

Ex(x, y=0, z)=0   [9]

implies that c4 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 c4 must be zero for the boundary condition in [9] to be satisfied everywhere it is required.

Next, the second boundary condition,

Ex(x, y=b, z)=0   [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, c3 could be zero. However, if c3 is zero (and we have already determined that c4 is zero), then all of the fields would end up being zero, because the function Y(y) in [7] would be zero everywhere. Hence, c3 cannot be zero if we are looking for a nonzero solution. Hence, the only alternative is if the above equation implies that:

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

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

Ey(x=0, y, z)=0

Ey(x=a, y, z)=0

Then (using identical reasoning to that above), we can determine that c2=0 and that:

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

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 Fz 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 c1, c3, and c6 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 TEmn, 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:

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

$$E_y = \frac{1}{\varepsilon}\frac{\partial F_z}{\partial x}$$

$$E_z = 0$$

$$H_x = -j\frac{1}{\omega\mu\varepsilon}\frac{\partial^2 F_z}{\partial x \partial z}$$

$$H_y = -j\frac{1}{\omega\mu\varepsilon}\frac{\partial^2 F_z}{\partial y \partial z}$$

$$H_z = -j\frac{1}{\omega\mu\varepsilon}\left(\frac{\partial^2}{\partial z^2}+k^2\right)F_z$$

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

$$E_x = A_{mn}\frac{n\pi}{b\varepsilon}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-jk_z z}$$

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

$$E_z = 0$$

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

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

$$H_z = -jA_{mn}\frac{k^2-k_z^2}{\omega\mu\varepsilon}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-jk_z z}$$

In the above, the constants are written as Amn - 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, Ex and Hy do not have independent coefficients).

Cutoff Frequency (fc)

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

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \omega^2 \mu \varepsilon = \left(\frac{2\pi f}{c}\right)^2$$

www.antenna-theory.com

[3]

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

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

[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 kz 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 kz 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:

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

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 (fc) 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 fc, 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 TEmn 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 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 B can be written as:

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

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

In a source free region, it can be shown that 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 Az component of the magnetic vector potential, via the following relationships:

$$E_x = -j \frac{1}{\omega\mu\varepsilon} \frac{\partial^2 A_z}{\partial x \partial z} \qquad\qquad H_x = \frac{1}{\mu} \frac{\partial A_z}{\partial y}$$

$$E_y = -j \frac{1}{\omega\mu\varepsilon} \frac{\partial^2 A_z}{\partial y \partial z} \qquad\qquad H_y = -\frac{1}{\mu} \frac{\partial A_z}{\partial x}$$

$$E_z = -j \frac{1}{\omega\mu\varepsilon} \left(\frac{\partial^2}{\partial z^2} + k^2 \right) A_z \qquad\qquad H_z = 0$$

To solve for Az (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 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 Az:

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

[1]

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

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

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

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

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

$$H_y = -B_{mn} \frac{k_x}{\mu} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$
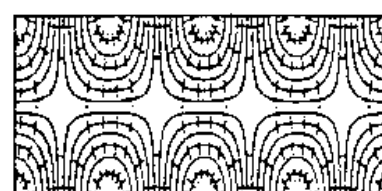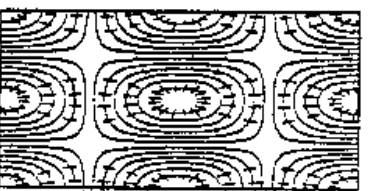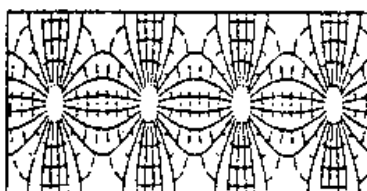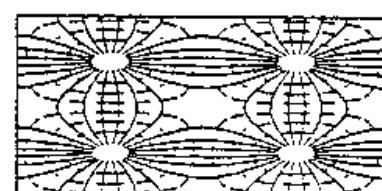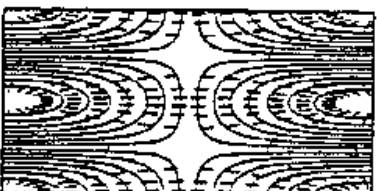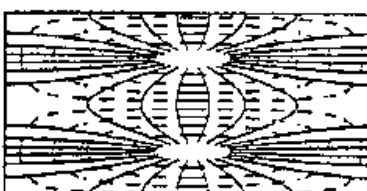
$$H_z = 0$$

In the above, k is again the wavenumber, and Bmn 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 TM0n and TMm0 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 Az 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 TMmn mode:

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

MODES

TE_{10} TE_{01} TE_{20}

TE_{11} TM_{11} TE_{21}

TM_{21} TE_{30} TE_{31}

TM_{31} TE_{40} TE_{02}

TE_{12} TM_{12} TE_{41}

TE_{22} TM_{41} TM_{22}

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 $I_s(m) = I_a(m)/10^4$	m	$\in \{1,...4\}$
dielectric tensor	$\overset{\leftrightarrow}{\varepsilon}$	$\overset{\leftrightarrow}{I}$
damping	δ_0	$10^{-2}...10^{-8}$
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 $\overrightarrow{e_{||}}$ 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 $P^{pla}(\omega_r; \omega_i)$ 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 $\omega_i > 0$ chosen so as to resolve the response peaks in $\Im m(P^{pla})$. 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 $(N_s = N_\theta = 8$ for LFEM, $N_s = N_\theta = 4$ for CFEM). All the Fourier modes $m \in \{0, ...N_\theta/2\}$ representable by the numerical discretization are excited with azimuthal currents $I_a(m)$ for TE modes, and axial currents $I_s(m)$ 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 $TE_{1,1}$ has here been calculated on a coarse homogeneous mesh $N_s = N_\theta = 6$. 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 $\vec{E}_\perp \sim \vec{A}_\perp = A_n \vec{e}_n + A_b \vec{e}_b$.
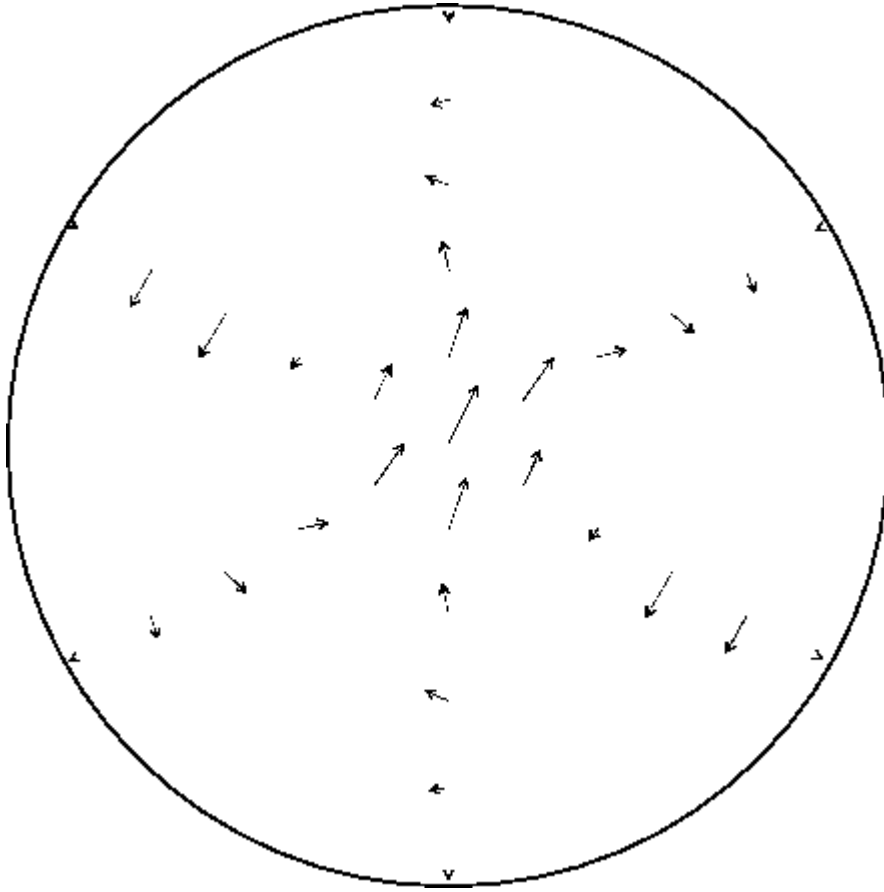
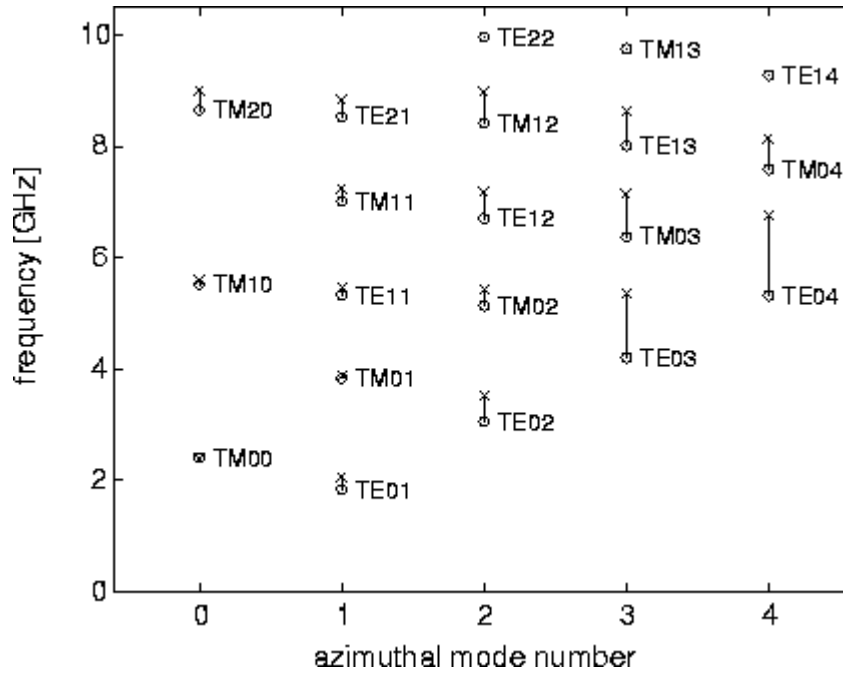Figure 5: Re(A_perp) for an eigenmode TE_11 calculated with CFEM.

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 $\Delta s \sim 1, 2, 4, 8, 16...$).

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.
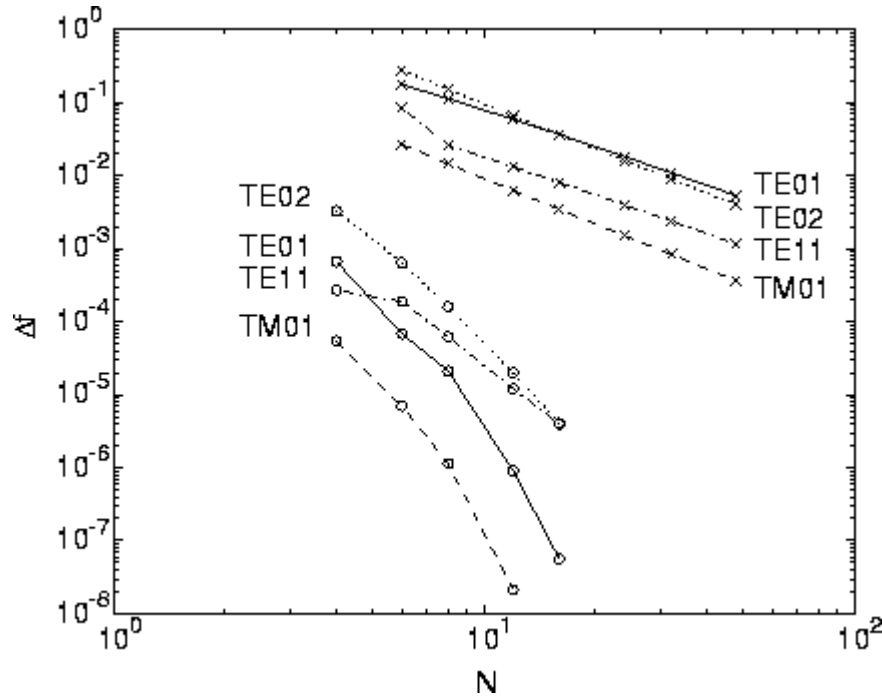
Figure 7: Convergence to the analytical result: relative frequency deviation Delta f versus the number of mesh intervals (N=N_s=N_theta) for the eigenmodes TE_01,TE_02,TE_11,TM_00} using LFEM (x-marks) and CFEM (circles).

Fig.7 shows the convergence of $\Delta f = (f^{num} - f^{th})/f^{th}$ 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.
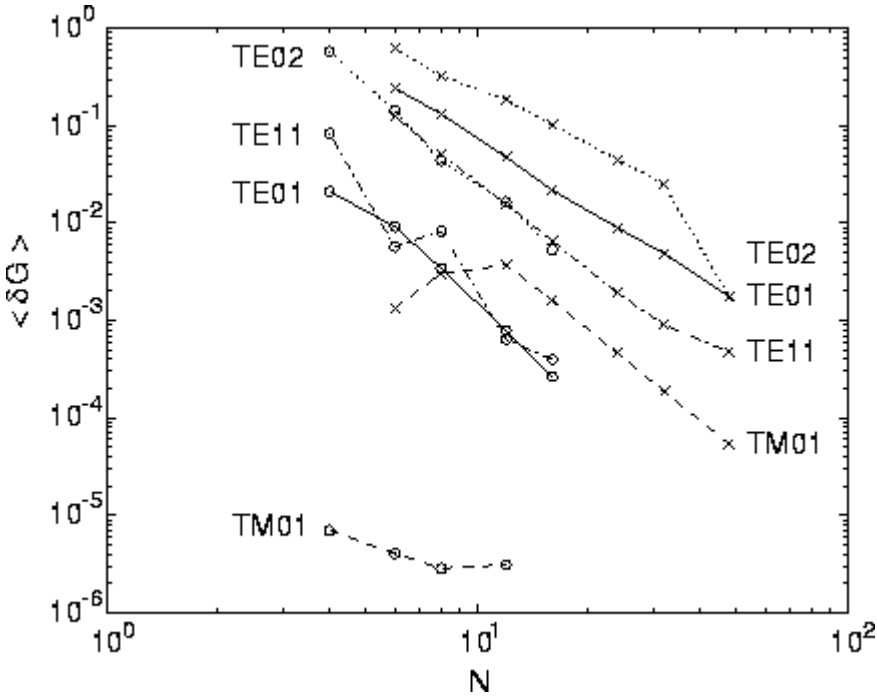
Figure 8: Precision of the gauge versus the number of mesh intervals (N=N_s=N_theta) for the eigenmodes TE_01,TE_02,TE_11,TM_00 using LFEM (x-marks) and CFEM (circles).

Convergence is also achieved for the gauge: fig.8 shows that the volume averaged gauge precision $\langle \delta G \rangle = \left\langle \dfrac{a|\nabla \cdot \vec{A}|}{|\vec{A}|} \right\rangle$ 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:

$$n \cdot (D_1 - D_2) = \tau, \tag{34}$$

$$n \wedge (E_1 - E_2) = 0, \tag{35}$$

$$n \cdot (B_1 - B_2) = 0, \tag{36}$$

$$n \wedge (H_1 - H_2) = K, \tag{37}$$

where τ is used for the surface change density (to avoid confusion with the conductivity), and K is the surface current density. Here, 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 ($\sigma \to \infty$) the tangential component of E vanishes, whereas that of H may remain finite. Let us examine the behaviour of the normal components.

Let medium 1 be a good conductor for which $\sigma / \varepsilon \varepsilon_0 \omega \gg 1$, whilst medium 2 is a perfect insulator. The surface change density is related to the currents flowing inside the conductor. In fact, the conservation of charge requires that

$$n \cdot j = \frac{\partial \tau}{\partial t} = -i\omega \tau. \tag{38}$$

However, $n \cdot j = n \cdot \sigma E_1$, so it follows from Eq. (6.1)(a) that

$$\left(1 + \frac{i\omega \varepsilon_0 \varepsilon_1}{\sigma}\right) n \cdot E_1 = \frac{i\omega \varepsilon_0 \varepsilon_2}{\sigma} n \cdot E_2. \tag{39}$$

It is clear that the normal component of E within the conductor also becomes vanishingly small as the conductivity approaches infinity.

If E vanishes inside a perfect conductor then the curl of E also vanishes, and the time rate of change of 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

$$n \cdot D = -\tau, \tag{40}$$

$$n \wedge E = 0, \tag{41}$$

$$n \cdot B = 0, \tag{42}$$

$$n \wedge H = -K. \tag{43}$$

Here, 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

$$H = \frac{1+i}{\sqrt{2}} \sqrt{\frac{\sigma}{\mu_0 \omega}} \, n \wedge E \tag{44}$$

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

$$|H_{||}| \simeq \frac{k}{\mu_0 \omega} |E_\perp| \qquad (45)$$

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

$$\frac{|E_{||}|}{|E_\perp|} \simeq k \sqrt{\frac{2}{\mu_0 \omega \sigma}} = \frac{d}{\lambda}, \qquad (46)$$

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

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0, \qquad (47)$$

where ψ represents any component of E or 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

$$E_x = E_1 \cos(k_1 x) \sin(k_2 y) \sin(k_3 z) e^{-i\omega t}, \tag{48}$$

$$E_y = E_2 \sin(k_1 x) \cos(k_2 y) \sin(k_3 z) e^{-i\omega t}, \tag{49}$$

$$E_z = E_3 \sin(k_1 x) \sin(k_2 y) \cos(k_3 z) e^{-i\omega t}, \tag{50}$$

where

$$k_1 = \frac{l\pi}{a}, \tag{51}$$

$$k_2 = \frac{m\pi}{b}, \tag{52}$$

$$k_3 = \frac{n\pi}{c}, \tag{53}$$

with l, m, n integers. The allowed frequencies are given by

$$\frac{\omega^2}{c^2} = \pi^2 \left(\frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{n^2}{c^2} \right). \tag{54}$$

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 $\nabla \wedge E = i\omega B$ 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 $\nabla \cdot E = 0$, which yields

$$k_1 E_1 + k_2 E_2 + k_3 E_3 = 0. \tag{55}$$

There are, in general, two linearly independent vectors 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 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, and 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 E at the walls can be estimated using Eq. (6.5):

$$E_{||} = \frac{1-i}{\sqrt{2}} \sqrt{\frac{\mu_0 \omega}{\sigma}} \, H_{||} \wedge n. \tag{56}$$

Now, the tangential component of 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

$$\overline{N} = \frac{1}{2}\operatorname{Re}(E_{\parallel} \wedge H_{\parallel}) = \frac{1}{2}\sqrt{\frac{\mu_0 \omega}{2\sigma}}\, H_{\parallel 0}^2\, n = \frac{H_{\parallel 0}^2}{2\sigma d}\, n, \tag{57}$$

where $H_{\parallel 0}$ 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), $H_{\parallel 0}$ is equal to the peak value of the surface current density K_0. It is helpful to define a surface resistance,

$$\overline{N} = \overline{K^2} R_s\, n, \tag{58}$$

where

$$R_s = \frac{1}{\sigma d}. \tag{59}$$

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

UNIT III

THE LINE AT RADIO FREQUENCY

There are two main forms of line at high frequency, namely

Open wire line

Coaxial line

At Radio Frequency G may be considered zero

Skin effect is considerable

Due to skin effect $\omega L \gg R$

Coaxial cable is used as a transmission line for radio frequency signals, in applications such as connecting radio transmitters and receivers with their antennas, computer network (Internet) connections, and distributing cable television signals. One advantage of coax over other types of transmission line is that in an ideal coaxial cable the electromagnetic field carrying the signal exists only in the space between the inner and outer conductors. This allows coaxial cable runs to be installed next to metal objects such as gutters without the power losses that occur in other transmission lines, and provides protection of the signal from external electromagnetic interference.

Coaxial cable differs from other shielded cable used for carrying lower frequency signals such as audio signals, in that the dimensions of the cable are controlled to produce a repeatable and predictable conductor spacing needed to function efficiently as a radio frequency transmission line.

How it works
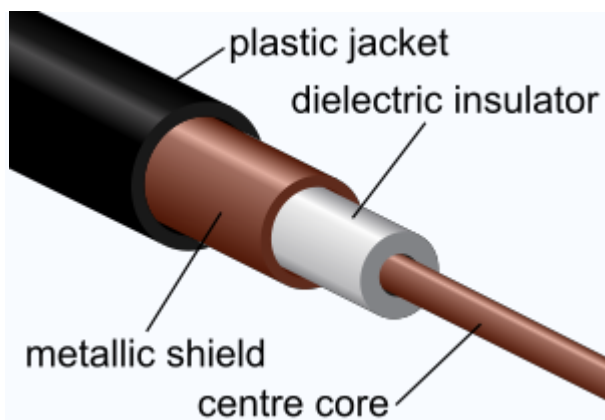
Coaxial cable cutaway

Like any electrical power cord, coaxial cable conducts AC electric current between locations. Like these other cables, it has two conductors, the central wire and the tubular shield. At any moment the current is traveling outward from the source in one of the conductors, and returning in the other. However, since it is alternating current, the current reverses direction many times a second. Coaxial cable differs from other cable because it is designed to carry radio frequency current. This has a

frequency much higher than the 50 or 60 Hz used in mains (electric power) cables, reversing direction millions to billions of times per second. Like other types of radio transmission line, this requires special construction to prevent power losses:

If an ordinary wire is used to carry high frequency currents, the wire acts as an antenna, and the high frequency currents radiate off the wire as radio waves, causing power losses. To prevent this, in coaxial cable one of the conductors is formed into a tube and encloses the other conductor. This confines the radio waves from the central conductor to the space inside the tube. To prevent the outer conductor, or shield, from radiating, it is connected to electrical ground, keeping it at a constant potential.

The dimensions and spacing of the conductors must be uniform. Any abrupt change in the spacing of the two conductors along the cable tends to reflect radio frequency power back toward the source, causing a condition called standing waves. This acts as a bottleneck, reducing the amount of power reaching the destination end of the cable. To hold the shield at a uniform distance from the central conductor, the space between the two is filled with a semirigid plastic dielectric. Manufacturers specify a minimum bend radius[2] to prevent kinks that would cause reflections. The connectors used with coax are designed to hold the correct spacing through the body of the connector.

Each type of coaxial cable has a characteristic impedance depending on its dimensions and materials used, which is the ratio of the voltage to the current in the cable. In order to prevent reflections at the destination end of the cable from causing standing waves, any equipment the cable is attached to must present an impedance equal to the characteristic impedance (called 'matching'). Thus the equipment "appears" electrically similar to a continuation of the cable, preventing reflections. Common values of characteristic impedance for coaxial cable are 50 and 75 ohms.

Description

Coaxial cable design choices affect physical size, frequency performance, attenuation, power handling capabilities, flexibility, strength and cost. The inner conductor might be solid or stranded; stranded is more flexible. To get better high-frequency performance, the inner conductor may be silver plated. Sometimes copper-plated iron wire is used as an inner conductor.

The insulator surrounding the inner conductor may be solid plastic, a foam plastic, or may be air with spacers supporting the inner wire. The properties of dielectric control some electrical properties of the cable. A common choice is a solid polyethylene (PE) insulator, used in lower-loss cables. Solid Teflon (PTFE) is also used as an insulator. Some coaxial lines use air (or some other gas) and have spacers to keep the inner conductor from touching the shield.

Many conventional coaxial cables use braided copper wire forming the shield. This allows the cable to be flexible, but it also means there are gaps in the shield layer, and the inner dimension of the shield varies slightly because the braid cannot be flat. Sometimes the braid is silver plated. For better shield performance, some cables have a double-layer shield. The shield might be just two braids, but it is more common now to have a thin foil shield covered by a wire braid. Some cables may invest in more than two shield layers, such as "quad-shield" which uses four alternating layers of foil and braid. Other shield designs sacrifice flexibility for better performance; some shields are a solid metal tube. Those cables cannot take sharp bends, as the shield will kink, causing losses in the cable.

For high power radio-frequency transmission up to about 1 GHz coaxial cable with a solid copper outer conductor is available in sizes of 0.25 inch upwards. The outer conductor is rippled like a bellows to permit flexibility and the inner conductor is held in position by a plastic spiral to approximate an air dielectric.

Coaxial cables require an internal structure of an insulating (dielectric) material to maintain the spacing between the center conductor and shield. The dielectric losses increase in this order: Ideal dielectric (no loss), vacuum, air, Polytetrafluoroethylene (PTFE), polyethylene foam, and solid polyethylene. A low relative permittivity allows for higher frequency usage. An inhomogeneous dielectric needs to be compensated by a non-circular conductor to avoid current hot-spots.

Most cables have a solid dielectric; others have a foam dielectric which contains as much air as possible to reduce the losses. Foam coax will have about 15% less attenuation but can absorb moisture—especially at its many surfaces—in humid environments, increasing the loss. Stars or spokes are even better but more expensive. Still more expensive were the air spaced coaxials used for some inter-city communications in the middle 20th Century. The center conductor was suspended by polyethylene discs every few centimeters. In a miniature coaxial cable such as an RG-62 type, the inner conductor is supported by a spiral strand of polyethylene, so that an air space exists between most of the conductor and the

inside of the jacket. The lower dielectric constant of air allows for a greater inner diameter at the same impedance and a greater outer diameter at the same cutoff frequency, lowering ohmic losses. Inner conductors are sometimes silver plated to smooth the surface and reduce losses due to skin effect. A rough surface prolongs the path for the current and concentrates the current at peaks and thus increases ohmic losses.

The insulating jacket can be made from many materials. A common choice is PVC, but some applications may require fire-resistant materials. Outdoor applications may require the jacket to resist ultraviolet light and oxidation. For internal chassis connections the insulating jacket may be omitted.

The ends of coaxial cables are usually made with RF connectors.

Open wire transmission lines have the property that the electromagnetic wave propagating down the line extends into the space surrounding the parallel wires. These lines have low loss, but also have undesirable characteristics. They cannot be bent, twisted or otherwise shaped without changing their characteristic impedance, causing reflection of the signal back toward the source. They also cannot be run along or attached to anything conductive, as the extended fields will induce currents in the nearby conductors causing unwanted radiation and detuning of the line. Coaxial lines solve this problem by confining the electromagnetic wave to the area inside the cable, between the center conductor and the shield. The transmission of energy in the line occurs totally through the dielectric inside the cable between the conductors. Coaxial lines can therefore be bent and moderately twisted without negative effects, and they can be strapped to conductive supports without inducing unwanted currents in them. In radio-frequency applications up to a few gigahertz, the wave propagates primarily in the transverse electric magnetic (TEM) mode, which means that the electric and magnetic fields are both perpendicular to the direction of propagation. However, above a certain cutoff frequency, transverse electric (TE) and/or transverse magnetic (TM) modes can also propagate, as they do in a waveguide. It is usually undesirable to transmit signals above the cutoff frequency, since it may cause multiple modes with different phase velocities to propagate, interfering with each other. The outer diameter is roughly inversely proportional to the cutoff frequency. A propagating surface-wave mode that does not involve or require the outer shield but only a single central conductor also exists in coax but this mode is effectively suppressed in coax of conventional geometry and common impedance. Electric field lines for

this TM mode have a longitudinal component and require line lengths of a half-wavelength or longer.

Connectors

A coaxial connector (male N-type).

Coaxial connectors are designed to maintain a coaxial form across the connection and have the same well-defined impedance as the attached cable. Connectors are often plated with high-conductivity metals such as silver or gold. Due to the skin effect, the RF signal is only carried by the plating and does not penetrate to the connector body. Although silver oxidizes quickly, the silver oxide that is produced is still conductive. While this may pose a cosmetic issue, it does not degrade performance.

Important parameters

Coaxial cable is a particular kind of transmission line, so the circuit models developed for general transmission lines are appropriate. See Telegrapher's equation.
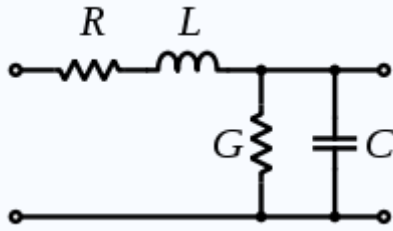
Schematic representation of the elementary components of a transmission line.
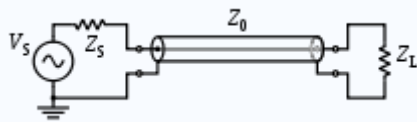
Schematic representation of a coaxial transmission line, showing the characteristic impedance Z_0.

Physical parameters

- Outside diameter of inner conductor, d.
- Inside diameter of the shield, D.
- Dielectric constant of the insulator, ε. The dielectric constant is often quoted as the relative dielectric constant ε_r referred to the dielectric constant of free space ε_0: $\varepsilon = \varepsilon_r \varepsilon_0$. When the insulator is a mixture of different dielectric materials (e.g., polyethylene foam is a mixture of polyethylene and air), then the term effective dielectric constant ε_{eff} is often used.
- Magnetic permeability of the insulator. μ Permeability is often quoted as the relative permeability μ_r referred to the permeability of free space μ_0: $\mu = \mu_r \mu_0$. The relative permeability will almost always be 1.

Fundamental electrical parameters

☐ Shunt capacitance per unit length, in farads per metre.

$$C = \frac{2\pi\epsilon}{\ln(D/d)} = \frac{2\pi\epsilon_0\epsilon_r}{\ln(D/d)}$$

☐ Series inductance per unit length, in henrys per metre.

$$L = \frac{\mu}{2\pi}\ln(D/d) = \frac{\mu_0\mu_r}{2\pi}\ln(D/d)$$

- Series resistance per unit length, in ohms per metre. The resistance per unit length is just the resistance of inner conductor and the shield at low frequencies. At higher frequencies, skin effect increases the effective resistance by confining the conduction to a thin layer of each conductor.
- Shunt conductance per unit length, in siemens per metre. The shunt conductance is usually very small because insulators with good dielectric properties are used (a very low loss tangent). At high frequencies, a dielectric can have a significant resistive loss.

Derived electrical parameters

- Characteristic impedance in ohms (Ω). Neglecting resistance per unit length for most coaxial cables, the characteristic impedance is determined from the capacitance per unit length (C) and the inductance per unit length (L). The simplified expression is ($Z_0 = \sqrt{L/C}$). Those parameters are determined from the ratio of the inner (d) and outer (D) diameters and the dielectric constant (ε). The characteristic impedance is given by[3]

$$Z_0 = \frac{1}{2\pi}\sqrt{\frac{\mu}{\epsilon}}\ln\frac{D}{d} \approx \frac{138\Omega}{\sqrt{\epsilon_r}}\log_{10}\frac{D}{d}$$

Assuming the dielectric properties of the material inside the cable do not vary appreciably over the operating range of the cable, this impedance is frequency independent above about five times the shield cutoff frequency. For typical coaxial cables, the shield cutoff frequency is 600 (RG-6A) to 2,000 Hz (RG-58C).[4]

- Attenuation (loss) per unit length, in decibels per meter. This is dependent on the loss in the dielectric material filling the cable, and resistive losses in the center conductor and outer shield. These losses are frequency dependent, the losses becoming higher as the frequency increases. Skin effect losses in the conductors can be reduced by increasing the diameter of the cable. A

cable with twice the diameter will have half the skin effect resistance. Ignoring dielectric and other losses, the larger cable would halve the dB/meter loss. In designing a system, engineers consider not only the loss in the cable, but also the loss in the connectors.

- Velocity of propagation, in meters per second. The velocity of propagation depends on the dielectric constant and permeability (which is usually 1).

$$v = \frac{1}{\sqrt{\epsilon\mu}} = \frac{c}{\sqrt{\epsilon_r \mu_r}}$$

- Cutoff frequency is determined by the possibility of exciting other propagation modes in the coaxial cable. The average circumference of the insulator is $\pi(D + d) / 2$. Make that length equal to 1 wavelength in the dielectric. The TE01 cutoff frequency is therefore

$$f_c = \frac{1}{\pi\left(\frac{D+d}{2}\right)\sqrt{\mu\epsilon}} = \frac{c}{\pi\left(\frac{D+d}{2}\right)\sqrt{\mu_r\epsilon_r}}.$$

☐ Peak Voltage

Significance of impedance

The best coaxial cable impedances in high-power, high-voltage, and low-attenuation applications were experimentally determined in 1929 at Bell Laboratories to be 30, 60, and 77 Ω respectively. For an air dielectric coaxial cable with a diameter of 10 mm the attenuation is lowest at 77 ohms when calculated for 10 GHz. The curve showing the power handling maxima at 30 ohms can be found here:

Consider the skin effect. The magnitude of an alternating current in a conductor decays exponentially with distance beneath the surface, with the depth of penetration being proportional to the square root of the resistivity. This means that in a shield of finite thickness, some small amount of current will still be flowing on the opposite surface of the conductor. With a perfect conductor (i.e., zero resistivity), all of the current would flow at the surface, with no penetration into and through the conductor. Real cables have a shield made of an imperfect, although usually very good, conductor, so there will always be some leakage.

The gaps or holes, allow some of the electromagnetic field to penetrate to the other side. For example, braided shields have many small gaps. The gaps are smaller when using a foil (solid metal) shield, but there is still a seam running the length of the cable. Foil becomes increasingly rigid with increasing thickness, so a thin foil layer is often surrounded by a layer of braided metal, which offers greater flexibility for a given cross-section.

This type of leakage can also occur at locations of poor contact between connectors at either end of the cable.

Nodes and Anti nodes :

At any point on the transmission line voltage or current value is zero called nodes.

At any point on the line voltage or current value is maximum called Antinodes

Reflection Coefficient:

The characteristic impedance of a transmission line, and that the tx line gives rise to forward and backward travelling voltage and current waves. We will use this information to determine the voltage reflection coefficient, which relates the amplitude of the forward travelling wave to the amplitude of the backward travelling wave.

To begin, consider the transmission line with characteristic impedance Z0 attached to a load with impedance ZL:
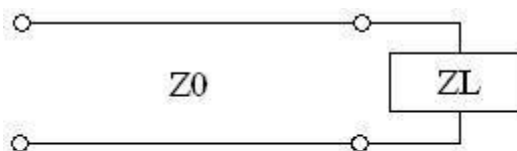
$$ZL = RL + jXL$$

At the terminals where the transmission line is connected to the load, the overall voltage must be given by:

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

[1]

Recall the expressions for the voltage and current on the line (derived on the previous page):

$$V(z,t) = V^+ e^{(j\omega t - \gamma z)} + V^- e^{(j\omega t + \gamma z)}$$

$$I(z,t) = I^+ e^{(j\omega t - \gamma z)} + I^- e^{(j\omega t + \gamma z)}$$

If we plug this into equation [1] (note that z is fixed, because we are evaluating this at a specific point, the end of the transmission line), we obtain:

$$\frac{V^+ + V^-}{I^+ - I^-} = \frac{V^+ + V^-}{V^+ - V^-} Z_0 = Z_L$$

The ratio of the reflected voltage amplitude to that of the forward voltage amplitude is the voltage reflection coefficient. This can be solved for via the above equation:

$$\Gamma = \frac{V^-}{V^+} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

The reflection coefficient is usually denoted by the symbol gamma. Note that the magnitude of the reflection coefficient does not depend on the length of the line, only the load impedance and the impedance of the transmission line. Also, note that if ZL=Z0, then the line is "matched". In this case, there is no mismatch loss and all power is transferred to the load. At this point, you should begin to understand the importance of impedance matching: grossly mismatched impedances will lead to most of the power reflected away from the load.

Note that the reflection coefficient can be a real or a complex number.

Standing Waves

We'll now look at standing waves on the transmission line. Assuming the propagation constant is purely imaginary (lossless line), We can re-write the voltage and current waves as:
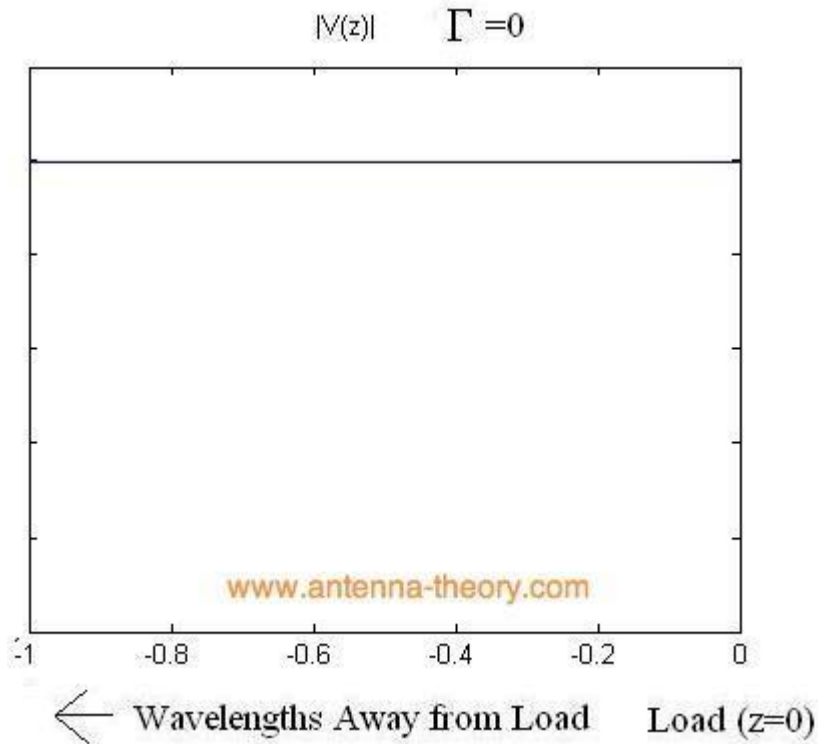
$$V(z) = V^{+}\left(e^{-j\beta z} + \Gamma e^{j\beta z}\right)$$

$$I(z) = \frac{V^{+}}{Z_{0}}\left(e^{-j\beta z} - \Gamma e^{j\beta z}\right)$$
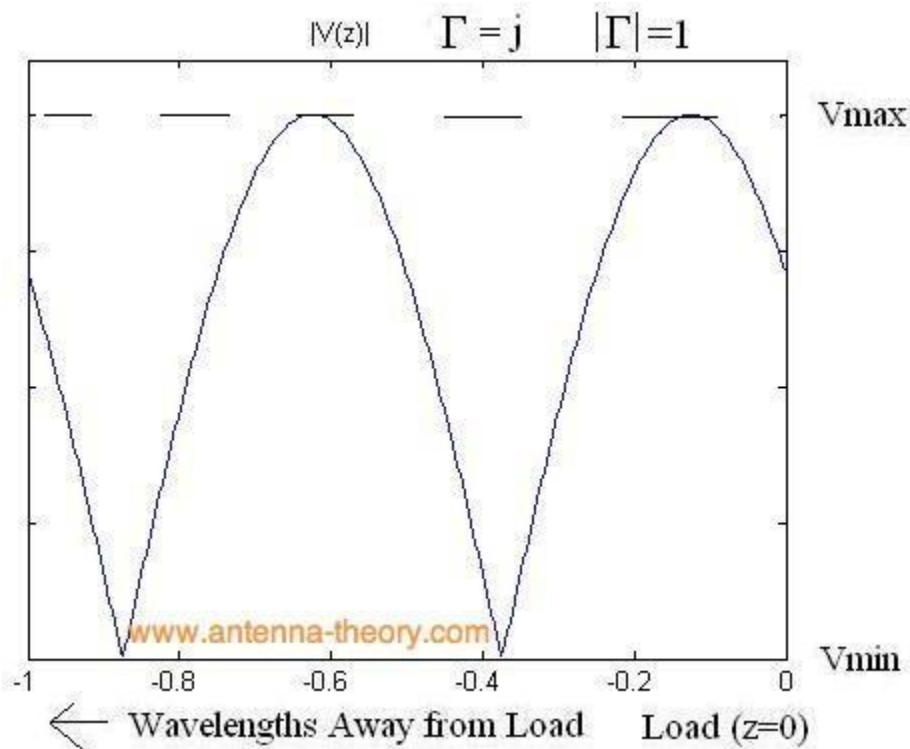
If we plot the voltage along the transmission line, we observe a series of peaks and minimums, which repeat a full cycle every half-wavelength. If gamma equals 0.5 (purely real), then the magnitude of the voltage would appear as:

Similarly, if gamma equals zero (no mismatch loss) the magnitude of the voltage would appear as:

$|V(z)|$ $\Gamma = 0$

-1   -0.8   -0.6   -0.4   -0.2   0

← Wavelengths Away from Load   Load (z=0)

Finally, if gamma has a magnitude of 1 (this occurs, for instance, if the load is entirely reactive while the transmission line has a Z0 that is real), then the magnitude of the voltage would appear as:

$|V(z)|$ $\quad \Gamma = j \quad |\Gamma| = 1$

Vmax

Vmin

← Wavelengths Away from Load   Load (z=0)

One thing that becomes obvious is that the ratio of Vmax to Vmin becomes larger as the reflection coefficient increases. That is, if the ratio of Vmax to Vmin is one, then there are no standing waves, and the impedance of the line is perfectly matched to the load. If the ratio of Vmax to Vmin is infinite, then the magnitude of the reflection coefficient is 1, so that all power is reflected. Hence, this ratio, known as the Voltage Standing Wave Ratio (VSWR) or standing wave ratio is a measure of how well matched a transmission line is to a load. It is defined as:

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

Input impedance of a transmission line:
  Determine the input impedance of a transmission line of length L attached to a load (antenna) with impedance ZA. Consider the following circuit:

In low frequency circuit theory, the input impedance would simply be ZA. However, for high-frequency (or long) transmission lines, we know that the voltage and the current are given by:

$$V(z) = V^+ \left(e^{-j\beta z} + \Gamma e^{j\beta z} \right)$$

$$I(z) = \frac{V^+}{Z_0} \left(e^{-j\beta z} - \Gamma e^{j\beta z} \right)$$

For simplicity, assume the transmission line is lossless, so that the propagation constant is purely imaginary. If we define z=0 to be at the terminals of the load or antenna, then we are interested in the ratio of the voltage to the current at location z=-L:

$$Z_{in}(z) = \frac{V(z)}{I(z)} = \frac{V^+ \left(e^{-j\beta z} + \Gamma e^{j\beta z} \right)}{V^+ \left(e^{-j\beta z} - \Gamma e^{j\beta z} \right)} Z_0$$

$$Z_{in}(z) = \left[\frac{e^{-j\beta z} + \Gamma e^{j\beta z}}{e^{-j\beta z} - \Gamma e^{j\beta z}} \right] Z_0$$

Using the definition for gamma (the voltage reflection coefficient), the above equation can be manipulated algebraically, and when evaluated at z=-L, we obtain:

$$Z_{in}(-L) = Z_0 \left[\frac{Z_A + jZ_0 \tan(\beta L)}{Z_0 + jZ_A \tan(\beta L)} \right]$$

This last equation is fundamnetal to understanding transmission lines. The input impedance of a load ZA is transformed by a transmission line as in the above equation. This equation can cause ZA to be transformed radically. An example will now be presented.
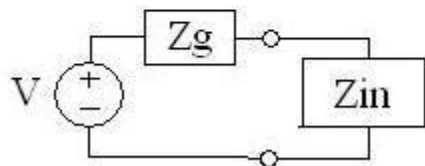
Example

Consider a voltage source, with generator impedance Zg, hooked to an antenna with impedance ZA via a transmission line. Suppose that Zg=50 Ohms, ZA=50 Ohms, Z0=200 Ohm, and that the line is a quarterwavelength long. How much power does the generator deliver?

Answer: The diagram for this problem is given in the following diagram:

(a) Actual Representation

(b) Equivalent Representation

The above diagram also shows the "equivalent circuit". The input impedance becomes:

$$\beta L = \left(\frac{2\pi}{\lambda}\right)\left(\frac{\lambda}{4}\right) = \frac{\pi}{2}$$

$$Z_{in}(-L) = Z_0\left[\frac{Z_A + jZ_0 \tan(\beta L)}{Z_0 + jZ_A \tan(\beta L)}\right]$$

$$= \left(200\left[\frac{50 + j(200)\tan\left(\frac{\pi}{2}\right)}{200 - j(50)\tan\left(\frac{\pi}{2}\right)}\right]\right) \rightarrow 200\frac{200}{50} = 800\Omega$$

Hence, the current that flows is given by:

$$I = \frac{V}{Z_g + Z_{in}} = \frac{V}{850}\ [Amps]$$

Note that if high frequency circuit theory was not taken into account, the current flow would have been V/100 Amps. This illustrates how transmission lines can upset the expected operation of high frequency circuits.

Quarter-Wave Transformer

Recall our formula for the input impedance of a transmission line of length L with characteristic impedance Z0 and connected to a load with impedance ZA:

$$Z_{in}(-L) = Z_0\left[\frac{Z_A + jZ_0 \tan(\beta L)}{Z_0 + jZ_A \tan(\beta L)}\right]$$

An interesting thing happens when the length of the line is a quarter of a wavelength:

$$Z_{in}\left(L=\frac{\lambda}{4}\right) = Z_0\left[\frac{Z_A + jZ_0\tan\left(\frac{2\pi}{\lambda}\frac{\lambda}{4}\right)}{Z_0 + jZ_A\tan\left(\frac{2\pi}{\lambda}\frac{\lambda}{4}\right)}\right]$$

$$\boxed{Z_{in}\left(L=\frac{\lambda}{4}\right) = \frac{Z_0^2}{Z_A}}$$

The above equation is important: it states that by using a quarter-wavelength of transmission line, the impedance of the load (ZA) can be transformed via the above equation. The utility of this operation can be seen via an example.

Example. Match a load with impedance ZA=100 Ohms to be 50 Ohms using a quarter-wave transformer, as shown below.

Solution: The problem is to determine Z0 (the characteristic impedance of our quarter-wavelength transmission line) such that the 100 Ohm load is matched to 50 Ohms. By applying the above equation, the problem is simple:

$$Z_{in}\left(L=\frac{\lambda}{4}\right) = \frac{Z_0^2}{Z_A}$$

$$50 = \frac{Z_0^2}{100} \Rightarrow Z_0 = \sqrt{(50)(100)} = 70.71\ \Omega$$

Hence, by using a transmission line with a characteristic impedance of 70.71 Ohms, the 100 Ohm load is matched to 50 Ohms. Hence, if a transmitter has an impedance of 50 Ohms and is trying to deliver power to the load (antenna), no power will be reflected back to the transmitter. In general, impedance matching is very important in RF/microwave circuit design. It is relatively simple at a single frequency, but becomes very difficult if wideband impedance matching is desired.

This technique is commonly employed with patch antennas. Circuits are printed as shown in the following figure. A 50 Ohm microstrip transmission line is matched to a patch antenna (impedance typically 200 Ohms or more) via a quarter-wavelength microstrip transmission line with the characteristic impedance chosen to match the load.

Because the quarter-wavelength transmission line is only a quarter-wavelength at a single frequency, this is a narrow-band matching technique. In the next section, we'll look at more uses of transmission lines.

Stub

In microwave and radio-frequency engineering, a stub is a length of transmission line or waveguide that is connected at one end only. The free end of the stub is either left open-circuit or (especially in the case of waveguides) short-circuited. Neglecting transmission line losses, the input impedance of the stub is purely reactive; either capacitive or inductive, depending on the electrical length of the

stub, and on whether it is open or short circuit. Stubs may thus be considered to be frequency-dependent capacitors and frequency-dependent inductors.

Because stubs take on reactive properties as a function of their electrical length, stubs are most common in UHF or microwave circuits where the line lengths are more manageable. Stubs are commonly used in antenna impedance matching circuits and frequency selective filters.

Smith charts can also be used to determine what length line to use to obtain a desired reactance.

Short circuited stub

The input impedance of a lossless short circuited line is,

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

where j is the imaginary unit, Z_0 is the characteristic impedance of the line, β is the phase constant of the line, and l is the physical length of the line.

Thus, depending on whether tan(βl) is positive or negative, the stub will be inductive or capacitive, respectively.

The Length of a stub to act as a capacitor C at an angular frequency of ω is then given by:

$$l = \frac{1}{\beta} \left[(n+1)\pi - \arctan\left(\frac{1}{\omega C Z_0} \right) \right]$$

The length of a stub to act as an inductor L at the same frequency is given by:

$$l = \frac{1}{\beta} \left[n\pi + \arctan\left(\frac{\omega L}{Z_0} \right) \right]$$

Open circuited stub

The input impedance of a lossless open circuit stub is given by

$$Z_{\text{OC}} = -jZ_0 \cot(\beta l)$$

It follows that whether $\cot(\beta l)$ is positive or negative, the stub will be capacitive or inductive, respectively.
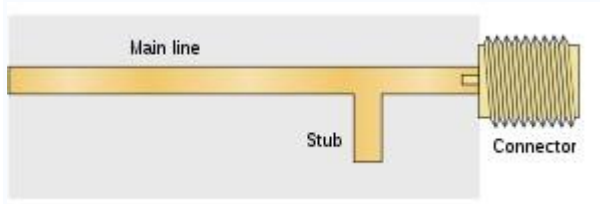
The length of an open circuit stub to act as an Inductor L at an angular frequency of ω is:

$$l = \frac{1}{\beta}\left[(n+1)\pi - \text{arccot}\left(\frac{\omega L}{Z_0}\right)\right]$$

The length of an open circuit stub to act as a capacitor C at the same frequency is:

$$l = \frac{1}{\beta}\left[n\pi + \text{arccot}\left(\frac{1}{\omega C Z_0}\right)\right]$$

Stub matching

In a stripline circuit, a stub may be placed just before an output connector to compensate for small mismatches due to the device's output load or the connector itself.

Stubs can be used to match a load impedance to the transmission line characteristic impedance. The stub is positioned a distance from the load. This distance is chosen so that at that point the resistive part of the load impedance is made equal to the resistive part of the characteristic impedance by impedance transformer action of the length of the main line. The length of the stub is chosen so that it exactly cancels the reactive part of the presented impedance. That is, the stub is made capacitive or inductive according to whether the main line is presenting an inductive or capacitive impedance respectively. This is not the same as the actual impedance of the load since the reactive part of the load impedance will be subject to impedance transformer action as well as the resistive part. Matching stubs can be made adjustable so that matching can be corrected on test.

single stub will only achieve a perfect match at one specific frequency. For wideband matching several stubs may be used spaced along the main transmission line. The resulting structure is filter-like and filter design techniques are applied. For instance, the matching network may be designed as a Chebyshev filter but is optimised for impedance matching instead of passband transmission. The resulting transmission function of the network has a passband ripple like the Chebyshev filter, but the ripples never reach 0dB insertion loss at any point in the passband, as they would do for the standard filter.

Single stub impedance matching:

- The load should be matched to the characteristic impedance of the line so that as much power as possible is transmitted from the generator to the load for radio-frequency power transmission.

- The lines should be matched because reflections from mismatched loads and junctions will result in echoes and will distort the information-carrying signal for information transmission.

- Short-circuited (instead of open-circuited) stubs are used for impedance-matching on transmission lines.

- Single-stub method for impedance matching : an arbitrary load impedance can be matched to a transmission line by placing a single short-circuited stub in parallel with the line at a suitable location (Fig.8-12).
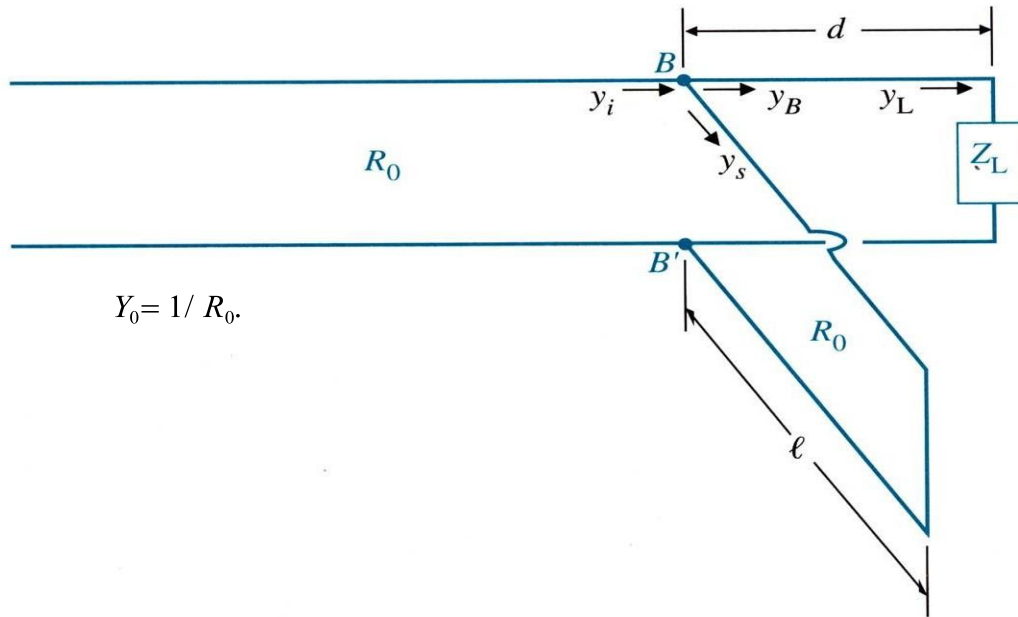
FIGURE 8-12 Impedance matching by single-stub method.

- $\quad$: the input admittance at BB' looking toward the load without the stub.

- $\quad$ Our impedance- (or admittance-) matching problem : to determine the location d and the length of the stub such that

$$Y_i = Y_0 = Y_B + Y_s, \tag{8-114}$$

Where Yo=1/Ro

- $\quad$ In terms of normalized admittance, the above equation becomes

$$1 = y_B + y_s, \tag{8-115}$$

- $\quad$ Ys purely imaginary (the input admittance of a short-circuited stub is purely susceptive).

- $\quad$ Thus, the above equation can be satisfied only if

$$y_B = 1 + jb_B \qquad\qquad (8\text{-}116)$$

and

$$y_s = -jb_B, \qquad\qquad (8\text{-}117)$$

where b_B can be either positive or negative.

- Our objectives : (1) to find the length d such that the admittance Y_B of the load section looking to the right of terminal BB' has a unitary real part, and (2) to find the length of the stub required to cancel the imaginary part.

Smith Chart:
The Smith Chart is a fantastic tool for visualizing the impedance of a transmission line and antenna system as a function of frequency. Smith Charts can be used to increase understanding of transmission lines and how they behave from an impedance viewpoint. Smith Charts are also extremely helpful for impedance matching, as we will see. The Smith Chart is used to display a real antenna's impedance when measured on a Vector Network Analyzer (VNA).

Smith Charts were originally developed around 1940 by Phillip Smith as a useful tool for making the equations involved in transmission lines easier to manipulate. See, for instance, the input impedance equation for a load attached to a transmission line of length L and characteristic impedance Z0. With modern computers, the Smith Chart is no longer used to the simplify the calculation of transmission line equatons; however, their value in visualizing the impedance of an antenna or a transmission line has not decreased.

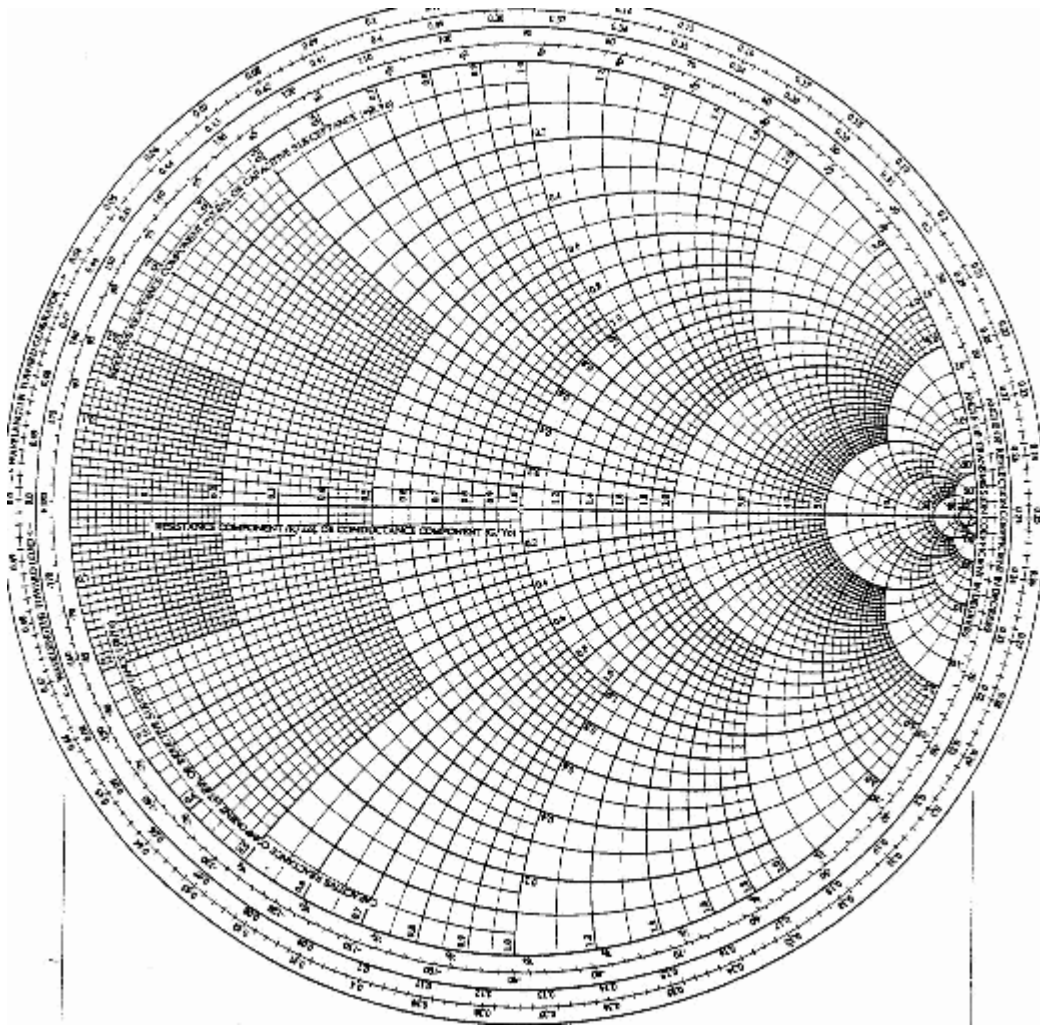The Smith Chart is shown in Figure 1. A larger version is shown here.

Figure 1. The basic Smith Chart.

Figure 1 should look a little intimidating, as it appears to be lines going everywhere. There is nothing to fear though. We will build up the Smith Chart from scratch, so that you can understand exactly what all of the lines mean. In fact, we are going to learn an even more complicated version of the Smith Chart known as the immitance Smith Chart, which is twice as complicated, but also twice as useful. But for now, just admire the Smith Chart and its curvy elegance.

This section of the antenna theory site will present an intro to the Smith Chart basics.

Smith Chart Tutorial

We'll now begin to explain the Smith Chart. The Smith Chart displays the complex reflection coefficient, in polar form, for an arbitrary impedance (we'll call the impedance ZL or the load impedance). For a primer on complex math, click here.

Recall that the complex reflection coefficient (Γ) for an impedance ZL attached to a transmission line with characteristic impedance Z0 is given by:

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}$$ [1]

For this tutorial, we will assume Z0 is 50 Ohms, which is often, but not always the case.

The complex reflection coefficient, or Γ , must have a magnitude between 0 and 1. As such, the set of all possible values for Γ must lie within the unit circle:

Figure 2. The Complex Reflection Coefficient must lie somewhere within the unit circle.

In Figure 2, we are plotting the set of all values for the complex reflection coefficient, along the real and imaginary axis. The center of the Smith Chart is the

point where the reflection coefficient is zero. That is, this is the only point on the smith chart where no power is reflected by the load impedance.

The outer ring of the Smith Chart is where the magnitude of Γ is equal to 1. This is the black circle in Figure 1. Along this curve, all of the power is reflected by the load impedance.

Normalized Load Impedance

To make the Smith Chart more general and independent of the characteristic impedance Z0 of the transmission line, we will normalize the load impedance ZL by Z0 for all future plots:

$$z_L = \frac{Z_L}{Z_0} = \frac{1+\Gamma}{1-\Gamma}$$
[1]

Equation [1] doesn't affect the reflection coefficient (Γ). It is just a convention that is used everywhere.

Constant Resistance Circles

for a given normalized load impedance zL, we can determine Γ and plot it on the Smith Chart.

Now, suppose we have the normalized load impedance given by:

$$z_1 = 1 + iY$$
[2]

In equation [2], Y is any real number. What would the curve corresponding to equation [2] look like if we plotted it on the Smith Chart for all values of Y? That is, if we plotted z1 = 1 + 0*i, and z1 = 1 + 10*i, z1 = 1 - 5*i, z1 = 1 - .333*i, ...
and any possible value for Y that you could think of, what is the resulting curve?
The answer is shown in Figure 1:

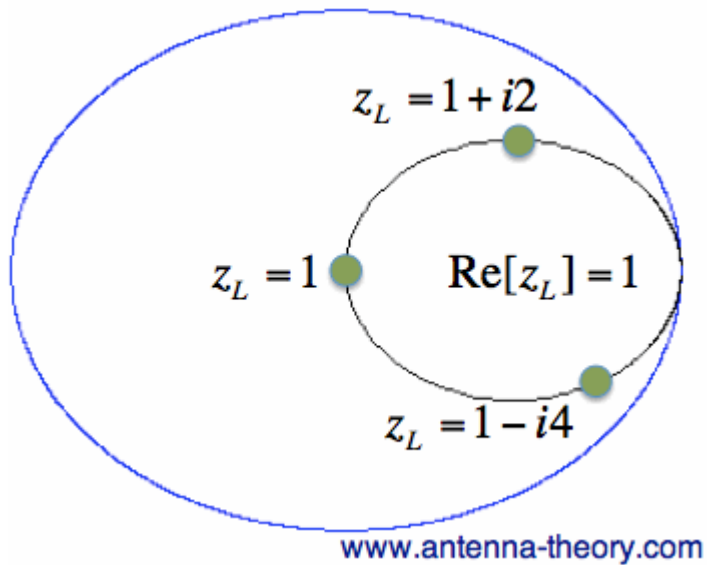Figure 1. Constant Resistance Circle for zL=1 on Smith Chart.

In Figure 1, the outer blue ring represents the boundary of the smith chart. The black curve is a constant resistance circle: this is where all values of z1 = 1 + i*Y will lie on. Several points are plotted along this curve, z1 = 1, z1 = 1 + i*2, and zL = 1 - i*4.

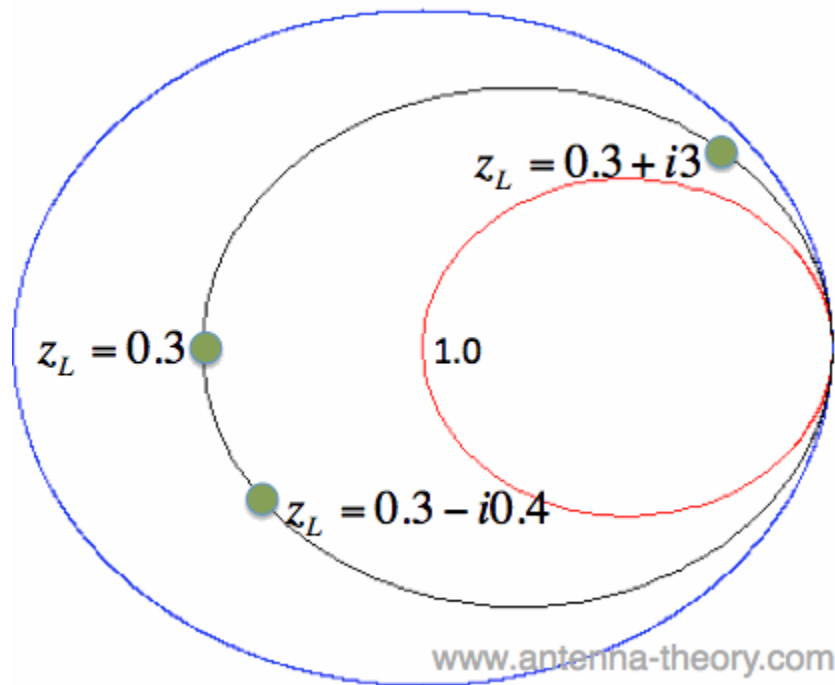Suppose we want to know what the curve z2 = 0.3 + i*Y looks like on the Smith Chart. The result is shown in Figure 2:

$z_L = 0.3 + i3$

$z_L = 0.3$

1.0

$z_L = 0.3 - i0.4$

www.antenna-theory.com

Figure 2. Constant Resistance Circle for zL=0.3 on Smith Chart.

In Figure 2, the black ring represents the set of all impedances where the real part of z2 equals 0.3. A few points along the circle are plotted. We've left the resistance circle of 1.0 in red on the Smith Chart.

These circles are called constant resistance curves. The real part of the load impedance is constant along each of these curves. We'll now add several values for the constant resistance, as shown in Figure 3:

Figure 3. Constant Resistance Circles on Smith Chart.

In Figure 3, the zL=0.1 resistance circle has been added in purple. The zL=6 resistance circle has been added in green, and zL=2 resistance circle is in black.

look at the set of curves defined by zL = R + iY, where Y is held constant and R varies from 0 to infinity. Since R cannot be negative for antennas or passive devices, we will restrict R to be greater than or equal to zero. As a first example, let zL = R + i. The curve defined by this set of impedances is shown in Figure 1:

Figure 1. Constant Reactance Curve for zL = R + i*1.

The resulting curve zL = R + i is plotted in green in Figure 1. A few points along the curve are illustrated as well. Observe that zL = 0.3 + i is at the intersection of the Re[zL] = 0.3 circle and the Im[zL]=1 curve. Similarly, observe that the zL = 2 + i point is at the intersection of the Re[zL]=2 circle and the Im[zL]=1 curve. (For a quick reminder of real and imaginary parts of complex numbers, see complex math primer.)

The constant reactance curve, defined by Im[zL]=-1 is shown in Figure 2:

Figure 2. Constant Reactance Curve for zL = R - i.

The resulting curve for Im[zL]=-1 is plotted in green in Figure 2. The point zL=1-i is placed on the Smith Chart, which is at the intersection of the Re[zL]=1 circle and the Im[zL]=-1 curve.

An important curve is given by Im[zL]=0. That is, the set of all impedances given by zL = R, where the imaginary part is zero and the real part (the resistance) is greater than or equal to zero. The result is shown in Figure 3:
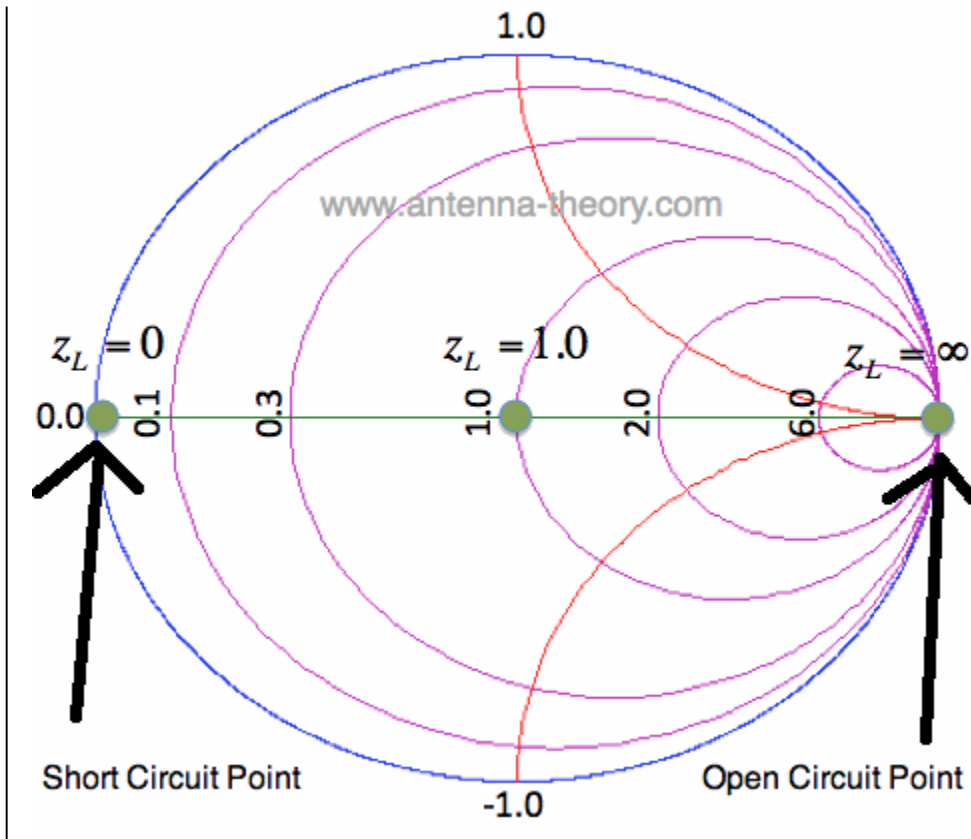
Figure 3. Constant Reactance Curve for zL=R.

The reactance curve given by Im[zL]=0 is a straight line across the Smith Chart. There are 3 special points along this curve. On the far left, where zL = 0 + i0, this is the point where the load is a short circuit, and thus the magnitude of Γ is 1, so all power is reflected.

In the center of the Smith Chart, we have the point given by zL = 1. At this location, Γ is 0, so the load is exactly matched to the transmission line. No power is reflected at this point.

The point on the far right in Figure 3 is given by zL = infinity. This is the open circuit location. Again, the magnitude of Γ is 1, so all power is reflected at this point, as expected.

Finally, we'll add a bunch of constant reactance curves on the Smith Chart, as shown in Figure 4.
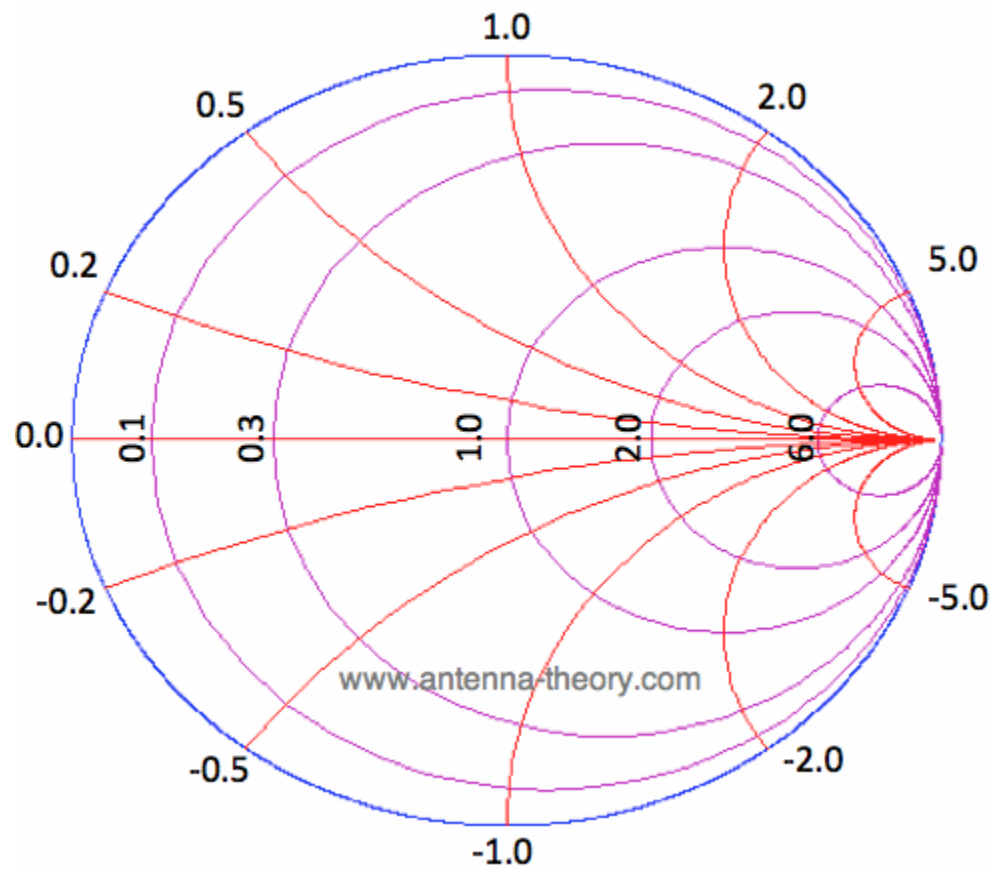
Figure 4. Smith Chart with Reactance Curves and Resistance Circles.

In Figure 4, we added constant reactance curves for Im[zL]=2, Im[zL]=5, Im[zL]=0.2, Im[zL]=0.5, Im[zL]=-2, Im[zL]=-5, Im[zL]=-0.2, and Im[zL] = -0.5.

Figure 4 shows the fundamental curves of the Smith

Chart. Applications of smith Chart:

Plotting an

impedance

Measurement

of VSWR

Measurement of reflection coefficient (magnitude and phase) Measurement of input impedance of the line

UNIT IV
GUIDED WAVES BETWEEN PARALLEL PLANES

The waves guided or directed by the guided structures are called guided wave. In general wave equations are derived from Maxwell's equation. To obtain the solution of this problem it is essential to apply certain restrictions or boundary conditions to the Maxwell's equation.

Maxwell's Equations are a set of four vector-differential equations that govern all of electromagnetics (except at the quantum level, in which case we as antenna people don't care so much). They were first presented in a complete form by James Clerk Maxwell back in the 1800s. He didn't come up with them all on his own, but did add the displacement current term to Ampere's law which made them complete.

The four equations (written only in terms of E and H, the electric field and the magnetic field), are given below.

$$\nabla \cdot \mathbf{E} = \frac{\rho_v}{\varepsilon} \qquad \text{(Gauss' Law)}$$

$$\nabla \cdot \mathbf{H} = 0 \qquad \text{(Gauss' Law for Magnetism)}$$

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \qquad \text{(Faraday's Law)}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Ampere's Law)}$$
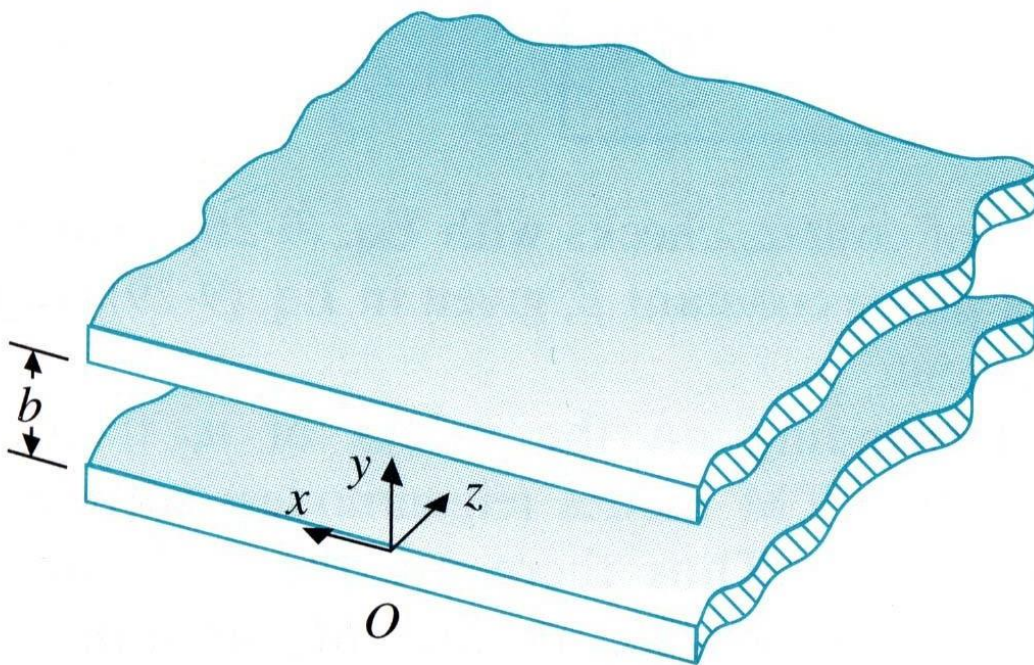
In Gauss' law, ρ_v is the volume electric charge density, J is the electric current density (in Amps/meter-squared), ε is the permittivity and μ is the permeability.

The good news about this is that all of electromagnetics is summed up in these 4 equations. The bad news is that no matter how good at math you are, these can only be solved with an analytical solution in extremely simple cases. Antennas don't present a very simple case, so these equations aren't used a whole lot in antenna theory (except for numerical methods, which numerically solve these approximately using a whole lot of computer power).

The last two equations (Faraday's law and Ampere's law) are responsible for electromagnetic radiation. The curl operator represents the spatial variation of the fields, which are coupled to the time variation. When the E-field travels, it is

altered in space, which gives rise to a time-varying magnetic field. A time-varying magnetic field then varies as a function of location (space), which gives rise to a time varying electric field. These equations wrap around each other in a sense, and give rise to a wave equation. These equations predict electromagnetic radiation as we understand it.

FIGURE 9-3   An infinite parallel-plate waveguide.

Consider a parallel-plate waveguide of two perfectly conducting plates separated by a distance b and filled with a dielectric medium having constitutive parameter as shown in Fig. 9-3. The plates are assumed to be infinite in extent in the x-direction. (Fields do not vary in the x-direction.)

   a) Obtain the time-harmonic field expressions for TM modes in the guide.

   b) Determine the cutoff frequency.

 a) For TM modes,

 - Eq. becomes

$$\frac{d^2 E_z^0(y)}{dy^2} + h^2 E_z^0(y) = 0. \qquad\qquad (9\text{-}44)$$

The general solution for Eq. (9-44) :
  - Boundary conditions (The tangential component of the electric field must vanish on the surface of the perfectly conducting plates.) :
    (i) At y=0 Ez=0
    (ii) At y=b Ez=0
  - The value of the eigenvalue h :

$$h = \frac{n\pi}{b}, \qquad n = 1, 2, 3, \dots . \qquad\qquad (9\text{-}46)$$

Types of propagation:
   TE waves
   TM waves
   TEM waves

Transverse electric(TE) wave has the magnetic field component in the direction of propagation, but no component of the electric field in the same direction. Hence the TE waves also known as M –waves or H-waves.

Transverse magnetic(TM) wave has the electric field in the direction of propagation, but no component of the magnetic field in the same direction. Hence the TM waves are also called E-waves.

Transverse electro magnetic(TEM) wave:
   No field in the direction of propagation

Attenuation of parallel plane guides:
    When the electromagnetic wave propagates through the wave guide, the amplitude of the fields or the signal strength of the wave decreases as the distance from the source increases. This is because when the wave strike the walls of the guide, the loss in the power takes place The attenuation factor is denoted by α.

A=Power lost per unit length/2*power transmitted

Attenuation due to finite wall conductivity is inversely proportional to the square root of wall conductivity, but depends on the mode and the frequency in a complicated way.

Attenuation due to wall losses in rectangular copper waveguide : Figure 9-7.

 TE10 mode has the lowest attenuation in a rectangular waveguide.

The attenuation constant increases rapidly toward infinity as the operating frequency approaches the cutoff frequency.

Causes for attenuation in waveguides : lossy dielectric and imperfectly conducting walls.

FIGURE 9-7 Attenuation due to wall losses in rectangular copper waveguide for TE_{10} and TM_{11} modes. $a = 2.29$ (cm), $b = 1.02$ (cm).

Cut-off frequency:

The frequency at which wave motion ceases is called cut-off frequency

$$f_c = \frac{n}{2b\sqrt{\mu\epsilon}} \quad \text{(Hz)}, \qquad (9\text{-}51)$$

Propagation Constant:

$$\gamma = \sqrt{\left(\frac{n\pi}{b}\right)^2 - \omega^2 \mu\epsilon}. \tag{9-50}$$

Wave impedance:

It is the ratio of the component of the electric field to that of magnetic field.
Wave impedance for TEM wave

$$Z_{TEM} = \frac{E_x^0}{H_y^0} = \frac{j\omega\mu}{\gamma_{TEM}} = \frac{\gamma_{TEM}}{j\omega\epsilon}, \tag{9-19}$$

Wave impedance for TM and TE wave

$$Z_{TM} = \frac{E_x^0}{H_y^0} = -\frac{E_y^0}{H_x^0} = \frac{\gamma}{j\omega\epsilon} \quad (\Omega). \tag{9-23}$$

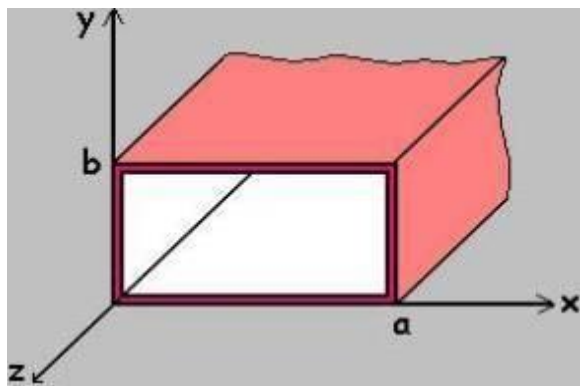$$Z_{TE} = \frac{\eta}{\sqrt{1 - (f_c/f)^2}} \quad (\Omega), \tag{9-39}$$

Phase velocity:

It is the velocity at which energy propagates along a wave guide

$$u_p = \frac{\omega}{\beta} \quad (m/s). \tag{7-58}$$

- The phase velocity and the wave impedance for TEM waves are independent of the frequency of the waves.
- TEM waves cannot exist in a single-conductor hollow (or dielectric-filled) waveguide of any shape.

Rectangular waveguides are the one of the earliest type of the transmission lines. They are used in many applications. A lot of components such as isolators, detectors, attenuators, couplers and slotted lines are available for various standard waveguide bands between 1 GHz to above 220 GHz.

A rectangular waveguide supports TM and TE modes but not TEM waves because we cannot define a unique voltage since there is only one conductor in a rectangular waveguide. The shape of a rectangular waveguide is as shown below. A material with permittivity e and permeability m fills the inside of the conductor.

A rectangular waveguide cannot propagate below some certain frequency. This frequency is called the cut-off frequency.

Here, we will discuss TM mode rectangular waveguides and TE mode rectangular waveguides separately. Let's start with the TM mode.

TM Modes

Consider the shape of the rectangular waveguide above with dimensions a and b (assume a>b) and the parameters e and m. For TM waves $H_z = 0$ and E_z should be solved from equation for TM mode;

$$\tilde{N}^2_{xy} E_z^0 + h^2 E_z^0 = 0$$

Since $E_z(x,y,z) = E_z^0(x,y)e^{-gz}$, we get the following equation,

$$(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + h^2)E_z^0(x,y) = 0$$

If we use the method of separation of variables, that is $E_z^0(x,y)=X(x).Y(y)$ we get,

$$-\frac{1}{X(x)}\frac{d^2 X(x)}{dx^2} = \frac{1}{Y(y)}\frac{d^2 Y(y)}{dy^2} + h^2$$

Since the right side contains x terms only and the left side contains y terms only, they are both equal to a constant. Calling that constant as k_x^2, we get;

$$\frac{d^2X(x)}{dx^2} + k_x^2 X(x) = 0$$

$$\frac{d^2Y(y)}{dy^2} + k_y^2 Y(y) = 0$$

where $k_y^2 = h^2 - k_x^2$

Now, we should solve for X and Y from the preceding equations. Also we have the boundary conditions of;

$E_z^0(0,y)=0$

$E_z^0(a,y)=0$

$E_z^0(x,0)=0$

$E_z^0(x,b)=0$

From all these, we conclude that

X(x) is in the form of sin $k_x x$, where $k_x=mp/a$, m=1,2,3,…

Y(y) is in the form of sin $k_y y$, where $k_y=np/b$, n=1,2,3,…

So the solution for $E_z^0(x,y)$ is

$$E_z^0(x,y) = E_0 \sin\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right) \text{ (V/m)}$$

From $k_y^2 = h^2 - k_x^2$, we have;

$$\boxed{h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

For TM waves, we have

$$H_x^0 = \frac{jw\varepsilon}{h^2}\frac{\partial E_z^0}{\partial y}$$

$$H_y^0 = -\frac{jw\varepsilon}{h^2}\frac{\partial E_z^0}{\partial x}$$

$$E_x^0 = -\frac{\gamma}{h^2}\frac{\partial E_z^0}{\partial x}$$

$$E_y^0 = -\frac{\gamma}{h^2}\frac{\partial E_z^0}{\partial y}$$

From these equations, we get

$$E_x^0(x,y) = -\frac{\gamma}{h^2}\left(\frac{m\pi}{a}\right)E_0 \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

$$\boxed{E_y^0(x,y) = -\frac{\gamma}{h^2}\left(\frac{n\pi}{b}\right)E_0 \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)}$$

$$\boxed{H_x^0(x,y) = \frac{jw\varepsilon}{h^2}\left(\frac{n\pi}{b}\right)E_0 \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)}$$

$$\boxed{H_y^0(x,y) = -\frac{jw\varepsilon}{h^2}\left(\frac{m\pi}{a}\right)E_0 \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)}$$

where

$$\boxed{\gamma = j\beta = j\sqrt{w^2\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}}$$

Here, m and n represent possible modes and it is designated as the TM$_{mn}$ mode. m denotes the number of half cycle variations of the fields in the x-direction and n denotes the number of half cycle variations of the fields in the y-direction.

When we observe the above equations we see that for TM modes in rectangular waveguides, neither m nor n can be zero. This is because of the fact that the field expressions are identically zero if either m or n is zero. Therefore, the lowest mode for rectangular waveguide TM mode is TM_{11} .

Here, the cut-off wave number is

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

and therefore,

$$\beta = \sqrt{k^2 - k_c^2}$$

The cut-off frequency is at the point where g vanishes. Therefore,

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}} \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \ (Hz)$$

Since l=u/f, we have the cut-off wavelength,

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

At a given operating frequency f, only those frequencies, which have $f_c < f$ will propagate. The modes with $f < f_c$ will lead to an imaginary b which means that the field components will decay exponentially and will not propagate. Such modes are called cut-off or evanescent modes.

The mode with the lowest cut-off frequency is called the dominant mode. Since TM modes for rectangular waveguides start from TM_{11} mode, the dominant frequency is

$$(f_c)_{11} = \frac{1}{2\sqrt{\varepsilon\mu}}\sqrt{\left(\frac{1}{a}\right)^2 + \left(\frac{1}{b}\right)^2} \ (Hz)$$

The wave impedance is defined as the ratio of the transverse electric and magnetic fields. Therefore, we get from the expressions for E_x and H_y (see the equations above);

$$Z_{TM} = \frac{E_x}{H_y} = \frac{\gamma}{jw\varepsilon} = \frac{j\beta}{jw\varepsilon} \Rightarrow Z_{TM} = \frac{\beta\eta}{k}$$

The guide wavelength is defined as the distance between two equal phase planes along the waveguide and it is equal to

$$\boxed{\lambda_g = \frac{2\pi}{\beta}} > \frac{2\pi}{k} = \lambda$$

which is thus greater than l, the wavelength of a plane wave in the filling medium.

The phase velocity is

$$u_p = \frac{w}{\beta} > \frac{w}{k} = \frac{1}{\sqrt{\mu\varepsilon}}$$
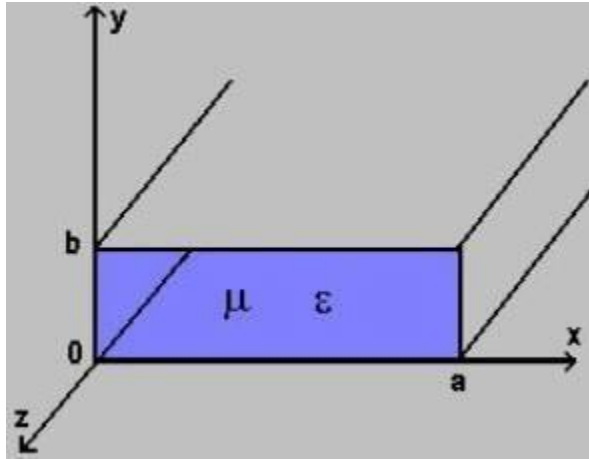
which is greater than the speed of light (plane wave) in the filling material.

Attenuation for propagating modes results when there are losses in the dielectric and in the imperfectly conducting guide walls. The attenuation constant due to the losses in the dielectric can be found as follows:

$$\gamma = j\beta = j\sqrt{k^2 - k_c^2} = jk\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu}\sqrt{\varepsilon + \frac{\sigma}{jw}}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

TE Modes

Consider again the rectangular waveguide below with dimensions a and b (assume a>b) and the parameters e and m.

For TE waves $E_z = 0$ and H_z should be solved from equation for TE mode;

$$\tilde{N}^2_{xy} H_z + h^2 H_z = 0$$

Since $H_z(x,y,z) = H^0_z(x,y)e^{-gz}$, we get the following equation,

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + h^2\right)H^0_z(x,y) = 0$$

If we use the method of separation of variables, that is $H_z^0(x,y) = X(x) \cdot Y(y)$ we get,

$$-\frac{1}{X(x)}\frac{d^2 X(x)}{dx^2} = \frac{1}{Y(y)}\frac{d^2 Y(y)}{dy^2} + h^2$$

Since the right side contains x terms only and the left side contains y terms only, they are both equal to a constant. Calling that constant as k_x^2, we get;

$$\frac{d^2 X(x)}{dx^2} + k_x^2 X(x) = 0$$

$$\frac{d^2 Y(y)}{dy^2} + k_y^2 Y(y) = 0$$

where $k_y^2 = h^2 - k_x^2$

Here, we must solve for X and Y from the preceding equations. Also we have the following boundary conditions:

$$\frac{\partial H^0_z}{\partial x} = 0 (E_y = 0) \quad \text{at x=0}$$

$$\frac{\partial H^0_z}{\partial x} = 0 (E_y = 0) \quad \text{at x=a}$$

$$\frac{\partial H^0_z}{\partial y} = 0 (E_x = 0) \quad \text{at y=0}$$

$$\frac{\partial H_z^0}{\partial y} = 0 (E_x = 0)$$

at y=b

From all these, we get

$$H_z^0(x,y) = H_0 \cos\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)$$ (A/m)

From $k_y^2 = h^2 - k_x^2$, we have;

$$\boxed{h^2 = \left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

For TE waves, we have

$$H_x^0 = -\frac{\gamma}{h^2}\frac{\partial H_z^0}{\partial x}$$

$$H_y^0 = -\frac{\gamma}{h^2}\frac{\partial H_z^0}{\partial y}$$

$$E_x^0 = -\frac{jw\mu}{h^2}\frac{\partial H_z^0}{\partial y}$$

$$E_y^0 = -\frac{jw\mu}{h^2}\frac{\partial H_z^0}{\partial x}$$

From these equations, we obtain

$$E_x^0(x,y) = \frac{jw\mu}{h^2}\left(\frac{n\pi}{b}\right)H_0 \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

$$\boxed{E_y^0(x,y) = -\frac{jw\mu}{h^2}\left(\frac{m\pi}{a}\right)H_0 \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)}$$

$$H_x^0(x,y) = \frac{\gamma}{h^2}\left(\frac{m\pi}{a}\right)H_0 \sin\left(\frac{m\pi}{a}x\right)\cos\left(\frac{n\pi}{b}y\right)$$

$$H_y^0(x,y) = \frac{\gamma}{h^2}\left(\frac{n\pi}{b}\right)H_0 \cos\left(\frac{m\pi}{a}x\right)\sin\left(\frac{n\pi}{b}y\right)$$

where

$$\gamma = j\beta = j\sqrt{w^2\mu\varepsilon - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2}$$

As explained before, m and n represent possible modes and it is shown as the TE$_{mn}$ mode. m denotes the number of half cycle variations of the fields in the x-direction and n denotes the number of half cycle variations of the fields in the y-direction.

Here, the cut-off wave number is

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

and therefore,

$$\beta = \sqrt{k^2 - k_c^2}$$

The cut-off frequency is at the point where g vanishes. Therefore,

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}\,(Hz)$$

Since l=u/f, we have the cut-off wavelength,

$$\lambda_c = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}}\,(m)$$

At a given operating frequency f, only those frequencies, which have f>f_c will propagate. The modes with f<f_c will not propagate.

The mode with the lowest cut-off frequency is called the dominant mode. Since TE$_{10}$ mode is the minimum possible mode that gives nonzero field expressions for rectangular waveguides, it is the dominant mode of a rectangular waveguide with a>b and so the dominant frequency is

$$(f_c)_{10} = \frac{1}{2a\sqrt{\mu\varepsilon}} \, (Hz)$$

The wave impedance is defined as the ratio of the transverse electric and magnetic fields. Therefore, we get from the expressions for E$_x$ and H$_y$ (see the equations above);

$$Z_{TE} = \frac{E_x}{H_y} = \frac{jw\mu}{\gamma} = \frac{jw\mu}{j\beta} \Rightarrow Z_{TE} = \frac{k\eta}{\beta}$$

The guide wavelength is defined as the distance between two equal phase planes along the waveguide and it is equal to

$$\boxed{\lambda_g = \frac{2\pi}{\beta}} > \frac{2\pi}{k} = \lambda$$

which is thus greater than l, the wavelength of a plane wave in the filling medium.

The phase velocity is

$$u_p = \frac{w}{\beta} > \frac{w}{k} = \frac{1}{\sqrt{\mu\varepsilon}}$$

which is greater than the speed of the plane wave in the filling material.

The attenuation constant due to the losses in the dielectric is obtained as follows:

$$\gamma = j\beta = j\sqrt{k^2 - k_c^2} = jk\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu\varepsilon}\sqrt{1 - \left(\frac{f_c}{f}\right)^2} = jw\sqrt{\mu}\sqrt{\varepsilon + \frac{\sigma}{jw}}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

After some manipulation, we get

$$\alpha_d = \frac{\sigma\eta}{2\sqrt{1-\left(\dfrac{f_c}{f}\right)^2}} = \frac{k^2 \tan\delta}{2\beta}$$

Example:

Consier a length of air-filled copper X-band waveguide, with dimensions a=2.286cm, b=1.016cm. Find the cut-off frequencies of the first four propagating modes.

Solution:

From the formula for the cut-off frequency

$$f_c = \frac{1}{2\sqrt{\varepsilon\mu}}\sqrt{\left(\frac{m}{a}\right)^2+\left(\frac{n}{b}\right)^2} \xrightarrow{air-filled} \frac{c}{2}\sqrt{\left(\frac{m}{a}\right)^2+\left(\frac{n}{b}\right)^2} \; (Hz)$$

(a)

(b)

(c)

Electric field lines
Magnetic field lines

FIGURE 9-5 Field lines for TE_{10} mode in rectangular waveguide.

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, fc. 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:

$$f_c = \frac{1}{2a\sqrt{\mu\varepsilon}} = \frac{c}{2a} \qquad \text{[Rectangular Cross Section]}$$

In the above, c is the speed of light within the waveguide, mu is the permeability of the material that fills the waveguide, and epsilon 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:

$$f_c = \frac{1.8412}{2\pi a\sqrt{\mu\varepsilon}} = \frac{1.8412c}{2\pi a} \qquad \text{[Circular Cross Section]}$$

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 Ez=0) or TM ('transverse magnetic' - which indicates that the H-field is orthogonal to the axis of the waveguide, so Hz = 0). The modes are further classified as TEij, 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 Ez and Hz are zero) mode. Their 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, 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 D and the corresponding electric field E as:

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

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

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 Hz and Ez are equal to zero. Hence, what field configurations that are allowed will be classified as either TM (Transverse Magnetic, in which Hz=0) and TE (Transverse Electric, in which Ez=0). The reason that waves cannot be TEM (Transverse Electromagnetic, Hz=Ez=0) will be shown towards the end of this derivation.

To perform our analysis, we'll assume that Ez=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:

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

$$E_y = \frac{1}{\varepsilon}\frac{\partial F_z}{\partial x}$$

$$E_z = 0$$

$$H_x = -j\frac{1}{\omega\mu\varepsilon}\frac{\partial^2 F_z}{\partial x \partial z}$$

$$H_y = -j\frac{1}{\omega\mu\varepsilon}\frac{\partial^2 F_z}{\partial y \partial z}$$

$$H_z = -j\frac{1}{\omega\mu\varepsilon}\left(\frac{\partial^2}{\partial z^2}+k^2\right)F_z$$

Therefore, if we can find Fz (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:

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

[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:

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

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

[2]

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

$$F_z(x, y, z) = X(x)Y(y)Z(z)$$
[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 Fz (equation [3]) into equation [2], and we end up with:

$$\frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} = -k^2$$
[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):

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

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

$$X'' + k_x^2 X = 0$$
$$Y'' + k_y^2 Y = 0$$
$$Z'' + k_z^2 Z = 0$$
[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:

$$X(x) = c_1 \cos(k_x x) + c_2 \sin(k_x x)$$
$$Y(y) = c_3 \cos(k_y y) + c_4 \sin(k_y y)$$
$$Z(z) = c_5 e^{jk_z z} + c_6 e^{-jk_z z}$$

[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 c5=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 Fz 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 Electric fields at any perfect conductor must be zero (why? because $J = \sigma E$, 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 Ex 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, Ey must be zero when x=0 and when x=a (independent of x and z). We will calculate Ex:

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

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

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

Ex(x, y=0, z)=0   [9]

implies that c4 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 c4 must be zero for the boundary condition in [9] to be satisfied everywhere it is required.

Next, the second boundary condition,

Ex(x, y=b, z)=0   [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, c3 could be zero. However, if c3 is zero (and we have already determined that c4 is zero), then all of the fields would end up being zero, because the function Y(y) in [7] would be zero everywhere. Hence, c3 cannot be zero if we are looking for a nonzero solution. Hence, the only alternative is if the above equation implies that:

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

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

Ey(x=0, y, z)=0

Ey(x=a, y, z)=0

Then (using identical reasoning to that above), we can determine that c2=0 and that:

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

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 Fz 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 c1, c3, and c6 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 TEmn, 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:

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

$$E_y = \frac{1}{\varepsilon}\frac{\partial F_z}{\partial x}$$

$$E_z = 0$$

$$H_x = -j\frac{1}{\omega\mu\varepsilon}\frac{\partial^2 F_z}{\partial x\partial z}$$

$$H_y = -j\frac{1}{\omega\mu\varepsilon}\frac{\partial^2 F_z}{\partial y\partial z}$$

$$H_z = -j\frac{1}{\omega\mu\varepsilon}\left(\frac{\partial^2}{\partial z^2}+k^2\right)F_z$$

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

$$E_x = A_{mn}\frac{n\pi}{b\varepsilon}\cos\left(\frac{m\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right)e^{-jk_z z}$$

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

$$E_z = 0$$

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

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

$$H_z = -jA_{mn}\frac{k^2-k_z^2}{\omega\mu\varepsilon}\cos\left(\frac{m\pi x}{a}\right)\cos\left(\frac{n\pi y}{b}\right)e^{-jk_z z}$$

In the above, the constants are written as Amn - 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, Ex and Hy do not have independent coefficients).

Cutoff Frequency (fc)

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

$$k_x^2 + k_y^2 + k_z^2 = k^2 = \omega^2 \mu \varepsilon = \left(\frac{2\pi f}{c} \right)^2$$

[3]

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

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

[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 kz 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 kz 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:

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

[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 (fc) 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 fc, 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 TEmn 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 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 B can be written as:

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

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

In a source free region, it can be shown that 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 Az component of the magnetic vector potential, via the following relationships:

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

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

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

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

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

$$H_z = 0$$

To solve for Az (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 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 Az:

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

[1]

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

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

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

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

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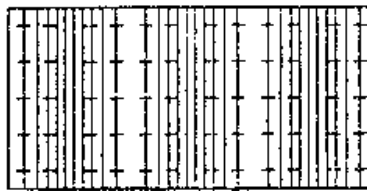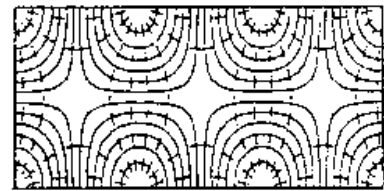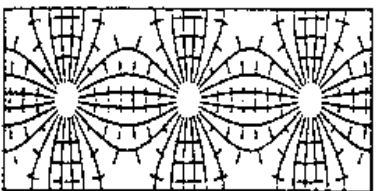
$$H_x = B_{mn} \frac{k_y}{\mu} \sin\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$

$$H_y = -B_{mn} \frac{k_x}{\mu} \cos\left(\frac{m\pi x}{a}\right) \sin\left(\frac{n\pi y}{b}\right) e^{-jk_z z}$$
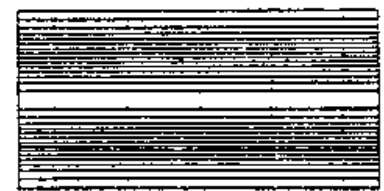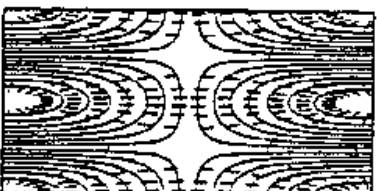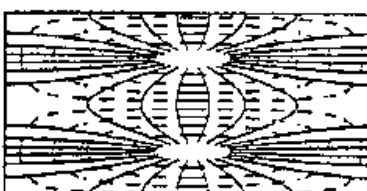
$$H_z = 0$$
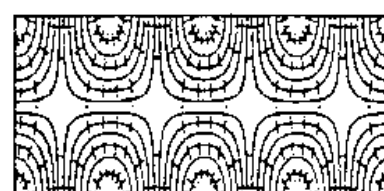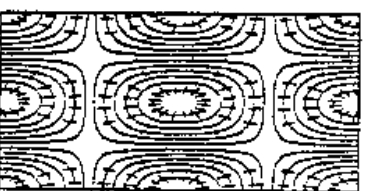
In the above, k is again the wavenumber, and Bmn is a constant, which determines the amplitude of the mn mode (a function of how much power is applied to the waveguide at that frequency).

Before discussing the modes, we must note that TM0n and TMm0 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 Az 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 TMmn mode:

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

MODES

TE_{10} TE_{01} TE_{20}

TE_{11} TM_{11} TE_{21}

TM_{21} TE_{30} TE_{31}

TM_{31} TE_{40} TE_{02}

TE_{12} TM_{12} TE_{41}

TE_{22} TM_{41} TM_{22}

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 $I_s(m) = I_a(m)/10^4$	m	$\in \{1,...4\}$
dielectric tensor	$\overleftrightarrow{\varepsilon}$	$\overleftrightarrow{I}$
damping	δ_0	$10^{-2}...10^{-8}$
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 $P^{pla}(\omega_r; \omega_i)$ 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 $\omega_i > 0$ chosen so as to resolve the response peaks in $\Im m(P^{pla})$. 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 $(N_s = N_\theta = 8$ for LFEM, $N_s = N_\theta = 4$ for CFEM). All the Fourier modes $m \in \{0, ...N_\theta/2\}$ representable by the numerical discretization are excited with azimuthal currents $I_a(m)$ for TE modes, and axial currents $I_s(m)$ 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 $TE_{1,1}$ has here been calculated on a coarse homogeneous mesh $N_s = N_\theta = 6$. 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 $\vec{E}_\perp \sim \vec{A}_\perp = A_n \vec{e}_n + A_b \vec{e}_b$.
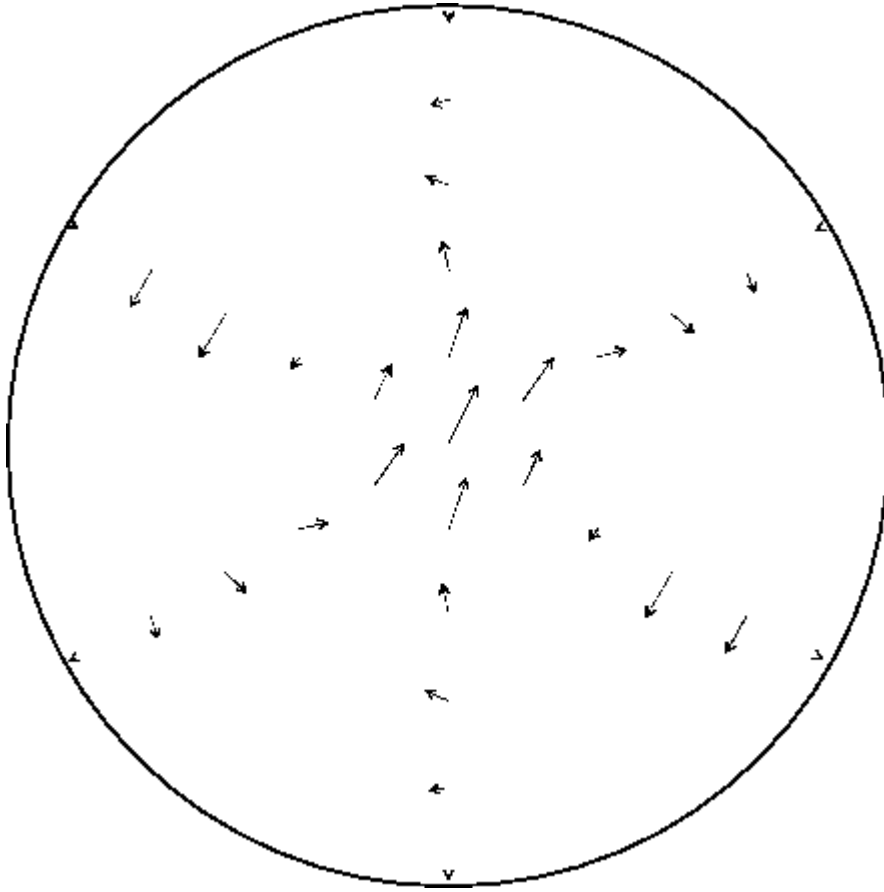
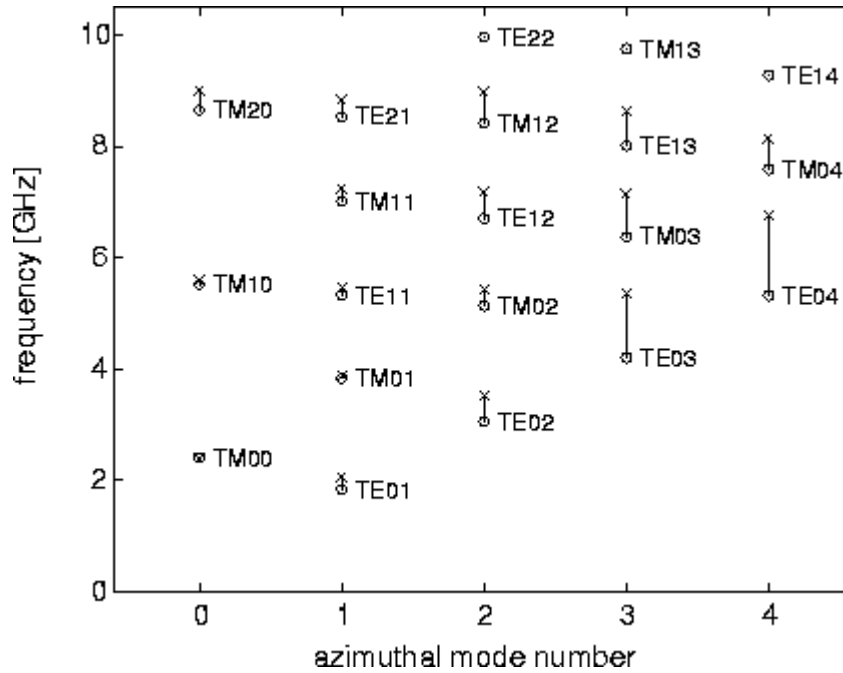Figure 5: Re(A_perp) for an eigenmode TE_11 calculated with CFEM.

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 $\Delta s \sim 1, 2, 4, 8, 16...$).

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.
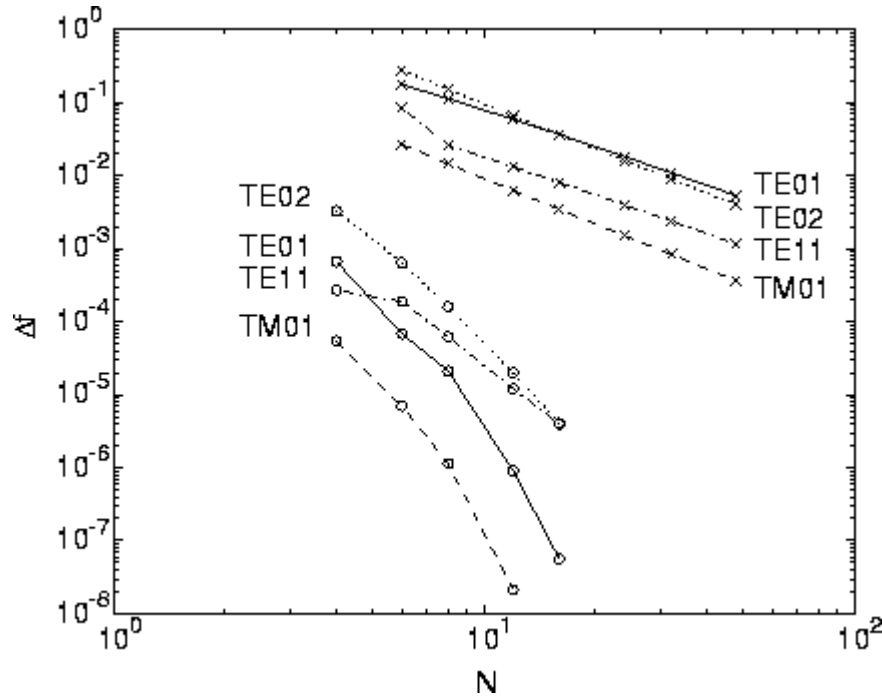
Figure 7: Convergence to the analytical result: relative frequency deviation Delta f versus the number of mesh intervals (N=N_s=N_theta) for the eigenmodes TE_01,TE_02,TE_11,TM_00} using LFEM (x-marks) and CFEM (circles).

Fig.7 shows the convergence of $\Delta f = (f^{num} - f^{th})/f^{th}$ 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.
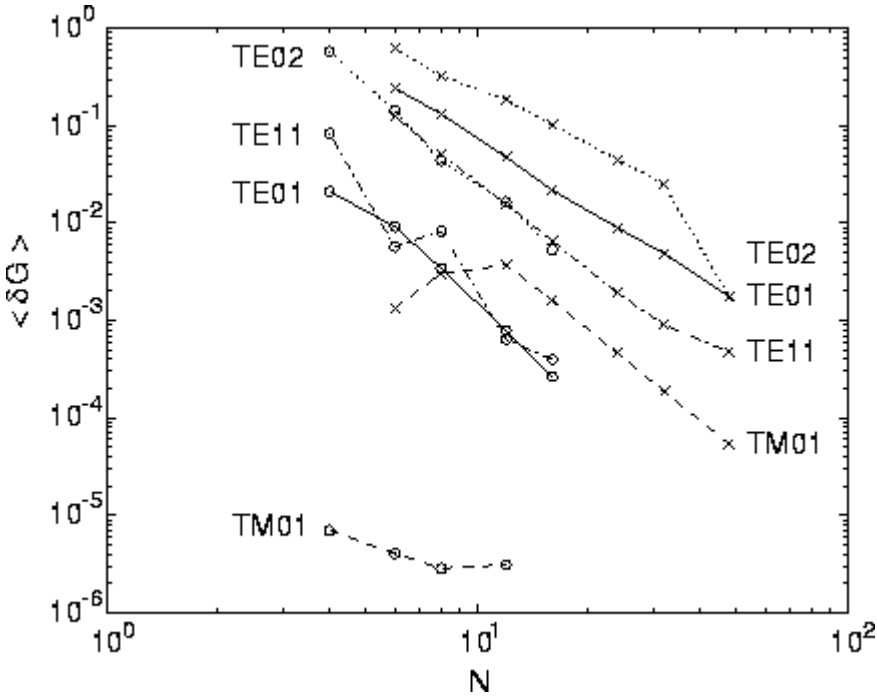
Figure 8: Precision of the gauge versus the number of mesh intervals (N=N_s=N_theta) for the eigenmodes TE_01,TE_02,TE_11,TM_00 using LFEM (x-marks) and CFEM (circles).

Convergence is also achieved for the gauge: fig.8 shows that the volume averaged gauge precision $\langle \delta G \rangle = \left\langle \frac{a|\nabla \cdot \vec{A}|}{|\vec{A}|} \right\rangle$ 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:

$$\boldsymbol{n}\cdot(\boldsymbol{D}_1 - \boldsymbol{D}_2) = \tau, \tag{34}$$

$$\boldsymbol{n}\wedge(\boldsymbol{E}_1 - \boldsymbol{E}_2) = 0, \tag{35}$$

$$\boldsymbol{n}\cdot(\boldsymbol{B}_1 - \boldsymbol{B}_2) = 0, \tag{36}$$

$$\boldsymbol{n}\wedge(\boldsymbol{H}_1 - \boldsymbol{H}_2) = \boldsymbol{K}, \tag{37}$$

where τ is used for the surface change density (to avoid confusion with the conductivity), and $\boldsymbol{K}$ is the surface current density. Here, $\boldsymbol{n}$ 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 ($\sigma \to \infty$) the tangential component of $\boldsymbol{E}$ vanishes, whereas that of $\boldsymbol{H}$ may remain finite. Let us examine the behaviour of the normal components.

Let medium 1 be a good conductor for which $\sigma/\epsilon\epsilon_0\omega \gg 1$, whilst medium 2 is a perfect insulator. The surface change density is related to the currents flowing inside the conductor. In fact, the conservation of charge requires that

$$\boldsymbol{n}\cdot\boldsymbol{j} = \frac{\partial\tau}{\partial t} = -i\omega\,\tau. \tag{38}$$

However, $\boldsymbol{n}\cdot\boldsymbol{j} = \boldsymbol{n}\cdot\sigma\boldsymbol{E}_1$, so it follows from Eq. (6.1)(a) that

$$\left(1 + \frac{i\omega\epsilon_0\epsilon_1}{\sigma}\right)\boldsymbol{n}\cdot\boldsymbol{E}_1 = \frac{i\omega\epsilon_0\epsilon_2}{\sigma}\,\boldsymbol{n}\cdot\boldsymbol{E}_2. \tag{39}$$

It is clear that the normal component of E within the conductor also becomes vanishingly small as the conductivity approaches infinity.

If E vanishes inside a perfect conductor then the curl of E also vanishes, and the time rate of change of 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

$$n \cdot D = -\tau, \qquad (40)$$

$$n \wedge E = 0, \qquad (41)$$

$$n \cdot B = 0, \qquad (42)$$

$$n \wedge H = -K. \qquad (43)$$

Here, 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

$$H = \frac{1+i}{\sqrt{2}} \sqrt{\frac{\sigma}{\mu_0 \omega}} \, n \wedge E \qquad (44)$$

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

H very nearly tangential and E very nearly normal is propagated along the surface of the metal. According to the Faraday-Maxwell equation

$$|H_\parallel| \simeq \frac{k}{\mu_0 \omega} |E_\perp| \qquad (45)$$

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 H is accompanied by a small tangential component of E. By comparing these two expressions, we obtain

$$\frac{|E_\parallel|}{|E_\perp|} \simeq k\sqrt{\frac{2}{\mu_0 \omega \sigma}} = \frac{d}{\lambda}, \qquad (46)$$

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 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 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

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0, \qquad (47)$$

where $\overset{\psi}{}$ represents any component of E or 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

$$E_x = E_1 \cos(k_1 x) \sin(k_2 y) \sin(k_3 z) e^{-i\omega t}, \tag{48}$$

$$E_y = E_2 \sin(k_1 x) \cos(k_2 y) \sin(k_3 z) e^{-i\omega t}, \tag{49}$$

$$E_z = E_3 \sin(k_1 x) \sin(k_2 y) \cos(k_3 z) e^{-i\omega t}, \tag{50}$$

where

$$k_1 = \frac{l\pi}{a}, \tag{51}$$

$$k_2 = \frac{m\pi}{b}, \tag{52}$$

$$k_3 = \frac{n\pi}{c}, \tag{53}$$

with l, m, n integers. The allowed frequencies are given by

$$\frac{\omega^2}{c^2} = \pi^2 \left(\frac{l^2}{a^2} + \frac{m^2}{b^2} + \frac{n^2}{c^2} \right). \tag{54}$$

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 $\nabla \wedge \boldsymbol{E} = i\omega \boldsymbol{B}$ 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 $\nabla \cdot \boldsymbol{E} = 0$, which yields

$$k_1 E_1 + k_2 E_2 + k_3 E_3 = 0. \tag{55}$$

There are, in general, two linearly independent vectors $\boldsymbol{E}$ 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 $\boldsymbol{E}$ 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, and 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 $\boldsymbol{E}$ at the walls can be estimated using Eq. (6.5):

$$E_{||} = \frac{1-i}{\sqrt{2}} \sqrt{\frac{\mu_0 \omega}{\sigma}} \, H_{||} \wedge n. \tag{56}$$

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

$$\overline{N} = \frac{1}{2} \operatorname{Re}(\mathbf{E}_{||} \wedge \mathbf{H}_{||}) = \frac{1}{2}\sqrt{\frac{\mu_0 \omega}{2\sigma}}\, H_{||0}^2\, \mathbf{n} = \frac{H_{||0}^2}{2\sigma d}\, \mathbf{n}, \tag{57}$$

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 the peak value of the surface current density K_0. It is helpful to define a surface resistance,

$$\overline{N} = \overline{K^2} R_s\, \mathbf{n}, \tag{58}$$

where

$$R_s = \frac{1}{\sigma d}. \tag{59}$$

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