Transmission Line Fundamentals

by

Chris Angove

DISCLAIMER

The Author makes no representation or warranties with respect to the accuracy or completeness of the contents of this paper and specifically disclaims any implied warranties of merchantability or fitness for any particular purpose and shall in no event be liable for any loss of profit or any other commercial damage, including but not limited to special, incidental, consequential or other damages. Any opinions expressed are those personal opinions of the Author only and they do not necessarily imply that any detailed and accredited engineering tests have been performed to arrive at them.

The reader is advised to check the information given, formulas and derivations against other sources before using them because the Author cannot guarantee them to be free from errors.

Accronym	Meaning
µm	micrometer
dB	decibel
DC	direct current
RF	radio frequency
TE	transverse electric
TEM	transverse electric magnetic
TM	transverse magnetic
VSWR	voltage standing wave ratio

Contents

Figures

1 TRANSMISSION LINES

1.1 What is an Electrical Transmission Line?

An electrical transmission line is a device for transferring electrical energy from one location to another. This must usually be achieved with the highest possible efficiency. Transmission lines have been developed for operation at practically all frequencies from DC to optical. There are three common types of transmission line used in radio frequency (RF) and microwave engineering:

- Transverse electric-magnetic (TEM) transmission lines.
- Conductor waveguides.
- Dielectric waveguides.

TEM transmission lines have separate conductors for the forward and return electric current paths and include open wire lines, twisted wire lines, coaxial cables and stripline. The common transmission line known as microstrip departs slightly from pure TEM as the field associated with the line is shared between the substrate material and the air above, so this line is sometimes called a quasi-TEM transmission line. A TEM transmission line will operate at frequencies from DC upwards. In practice however, there is an upper frequency limit determined by when the wavelength becomes sufficiently short that non-TEM modes start to occur. These are known as waveguide modes and comprise either transverse electric (TE) or transverse magnetic (TM) modes.

Conductor waveguides take the form of precision rigid pipes comprising insulator cores bound by electrical conductors. They allow propagation of electrical energy in the form of non-TEM wave modes supported by electrical currents circulating on the inside walls of the guide and the resulting magnetic and electric fields. The modes are again specific orders of TE or TM modes.

Dielectric waveguides comprise cores of low loss dielectric material surrounded by a shell of another similar dielectric material but with a slightly higher dielectric constant. Transmission occurs by a series of reflections at the dielectric boundaries. The most common example of the dielectric waveguide is the fiber optic cable, typically used for the propagation of infrared wavelengths around 1.5 μm

1.2 History

The theory of transmission lines developed from work performed by James Clerk Maxwell, Lord Kelvin and Oliver Heaviside. In 1855 Lord Kelvin performed the first distributed analysis of a transmission line [1]. He modelled the type of pulsed current then used in long distance telegraph cables and correctly predicted the poor performance of the trans-Atlantic submarine cable which was laid in 1858. In 1885 Heaviside published the first papers which analysed the propagation of telegraph-like signals, arriving at the telegrapher’s equations. These will be described in Section 1.3

Early telegraph lines were very crude, each one comprising a single *iron* conductor carried by overhead telegraph poles and forming an electrical circuit using an earth return path. The choice of an electrical conductor with such a relatively high resistivity would seem odd today, but in the mid-nineteenth century very little research had been done is this area and iron was plentiful and relatively cheap. After the invention of the telephone at the end of the nineteenth century, attempts to transmit reasonable quality audio frequencies (known as telephony) over telegraph lines were not successful. Two separate wires were found to be much more effective, the second providing the current return path instead of a route through the ground. It was soon discovered that using copper wires, a much better electrical conductor than iron, reduced the series loss affecting the signal. Heaviside also showed that the addition of series inductors, regularly spaced every mile or so, compensated for the capacitance of the line, increased its effectiveness and allowed a finer gauge of wire to be used between the inductors than previously.

1.3 The Elements of a Practical Uniform TEM Transmission Line

Even today the principle of the transmission line has not changed significantly from Heaviside's time. It comprises elements of serial capacitance, inductance, resistance and conductance distributed as evenly as possible along the line. The resistance and conductance require minimising as they cause unwanted attenuation of the signal. The capacitance and inductance require careful control to 'balance' each other out as much as possible.

A practical TEM transmission line will contain elements of series resistance R, series inductance L, parallel conductance G and parallel capacitance C distributed along the line. Consider a small element of such a transmission line of short elementary length δz, as shown in Figure 1-1. The lower case z represents the distance measured along the line which is also the direction of propagation. Figure 1-1 represents the section from an unbalanced transmission line, such as a coaxial cable which is the most common form of transmission line. In unbalanced transmission lines the 'return' path comprises a metallic screen surrounding the inner coaxial conductor used for the forward path. In this case the return path will have low resistance and inductance, which is assumed negligible and therefore not shown in Figure 1-1.

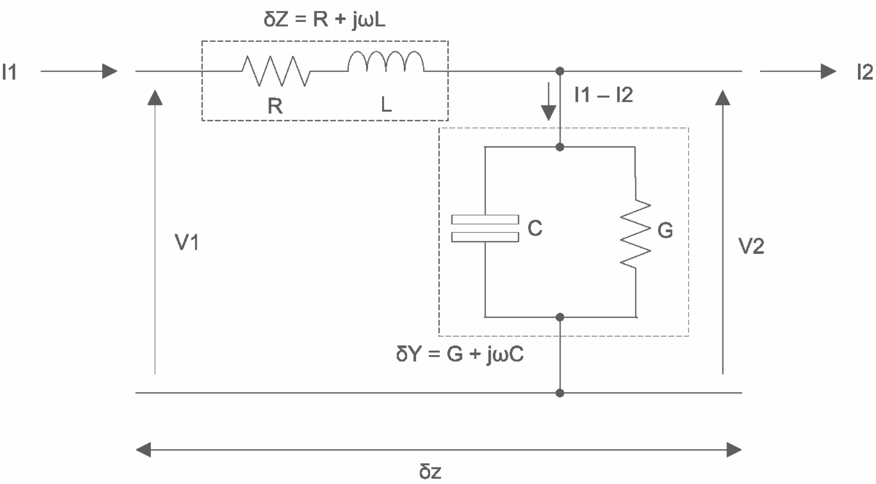

Figure 1-1 The fundamental element section of a uniform TEM transmission line

Many such elementary sections may be cascaded to form a section of a practical, uniform transmission line. The units of the elementary parameters are defined per unit length as follows:

- 'R' Ohms per metre (Ω / m);
- 'L' Henries per metre (H / m);
- 'G' Siemens per metre (S / m);
- 'C' Farads per metre (F / m);

The impedance of the series section represented by δZ, with an upper case Z, not to be confused with the length which is a lower case z is:

$$\delta Z = R + j\omega L \qquad (1.1)$$

and the admittance of the section is represented by δY, where

$$\delta Y = G + j\omega C \qquad (1.2)$$

In (1.1) and (1.2) the term ω is the angular frequency of the applied sinusoidal waveform in radians per second (rad/s). ω is related to the frequency in hertz (Hz) of the waveform by:

$$\omega = 2\pi f \qquad (1.3)$$

By Kirchoff's laws using the definitions of voltages and current shown in Figure 1-1:

$$V_2 = V_1 - I_1 Z \delta z \qquad (1.4)$$

$$I_2 = I_1 - V_2 Y \delta z \qquad (1.5)$$

If the voltage difference therefore across δz is δV

$$\delta V = V_2 - V_1 = -I_1 Z \delta z \qquad (1.6)$$

$$\begin{aligned} \frac{\delta V}{\delta z} &= -I_1 Z \\ \frac{\partial V}{\partial z} &= \left(\frac{\delta V}{\delta z}\right)_{\delta z \to 0}^{\lim} = -IZ \end{aligned} \qquad (1.7)$$

Similarly

$$\frac{\partial I}{\partial z} = -YV \qquad (1.8)$$

Differentiating (1.7) with respect to z:

$$\begin{aligned} \frac{\partial^2 V}{\partial z^2} &= -Z\frac{\partial I}{\partial z} = -Z(-YV) = YZV \\ \frac{\partial^2 V}{\partial z^2} &- YZV = 0 \end{aligned} \qquad (1.9)$$

This differential equation expresses the voltage variation along the line in terms of position. Similarly, the following differential equation may be derived in terms of the current.

$$\frac{\partial^2 I}{\partial z^2} - YZI = 0 \qquad (1.10)$$

1.4 The Total Voltage Wave

A practical linear transmission line will simply comprise a cascade of networks of the type shown in Figure 1-1. It will have some loss because finite values for resistance R and conductance G. Each of these parameters will dissipate heat and therefore waste some of the power intended for propagation along the transmission line. Each will also change with frequency, the resistance and conductance tending to increase with increasing frequency because of *the skin effect*. Excessive heat dissipation is not usually a problem at low power levels but may become so at high power levels if, for example, the transmission line was used to feed an antenna from a high power transmitter. Inductance also changes with frequency, tending to increase with increasing frequency.

Using a transmission line at high powers also increases the risk of breakdown and/or overheating. Breakdown can occur very quickly, often being initiated by a very narrow pulse of high instantaneous peak power. Overheating is normally the direct result of excessive *mean power* being dissipated in the R and G elements. Either can cause permanent damage to the transmission line.

The series element can be represented as an impedance Z, given by:

$$Z = R + j\omega L \tag{1.11}$$

and the parallel element can be represented by an admittance Y given by:

$$Y = G + j\omega C \tag{1.12}$$

(1.9) is a differential equation which has an exponential solution of the type

$$V = Ae^{-\gamma z} + Be^{\gamma z} \tag{1.13}$$

where A and B are constants and γ is the complex quantity given by

$$\gamma = \sqrt{ZY} = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{1.14}$$

Furthermore γ is known as the propagation constant and is given by

$$\gamma = \alpha + j\beta \tag{1.15}$$

where α is the attenuation constant expressed in Nepers per metre (Np/m) and β is the phase constant expressed in radians per metre (rad/m).

A Neper is a logarithmic way of expressing a power ratio, with a similar definition to the decibel, but using a natural logarithm instead of a common logarithm. The bases of logarithms may be changed using the following equation:

$$\ln b = \frac{\log_{10} b}{\log_{10} e} \tag{1.16}$$

Thus (1.16) may be used to convert Nepers to decibels.

(1.13) is the expression for the total voltage on the transmission line, the sum of the forward (V_F) and reverse (V_R) waves. Therefore

$$V_F = Ae^{-\gamma z} \tag{1.17}$$

$$V_R = Be^{\gamma z} \tag{1.18}$$

$$V = V_F + V_R \tag{1.19}$$

1.5 The Total Current Wave

Differentiating the total voltage from (1.13) with respect to z gives:

$$\frac{\partial V}{\partial z} = -\gamma Ae^{-\gamma z} + \gamma Be^{\gamma z} \tag{1.20}$$

But

$$\frac{\partial V}{\partial z} = -IZ \tag{1.21}$$

Therefore, using (1.14), (1.20) and (1.21), the total current waveform is given by:

$$I = \frac{Ae^{-\gamma z}}{\sqrt{\frac{Z}{Y}}} - \frac{Be^{-\gamma z}}{\sqrt{\frac{Z}{Y}}} \tag{1.22}$$

1.6 Characteristic Impedance

Another way of writing the total current waveform in terms of V_F and V_R is:

$$IZ_0 = V_F - V_R \tag{1.23}$$

where Z_0 is known as the characteristic impedance of the transmission line and is defined as:

$$Z_0 = \sqrt{\frac{Z}{Y}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{1.24}$$

As Z_0 is an impedance it has dimensions of Ohms (Ω). For some transmission lines operating at modest radio frequencies, R and G are negligible compared to the magnitudes of $j\omega L$ and $j\omega C$, so

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \tag{1.25}$$

The characteristic impedance is a basic property of the line determined by its physical construction and it describes the fundamental electrical matching capability of the transmission line. For a transmission line of infinite length, the impedance measured at one end will be Z_0. If the transmission line is of finite length and terminated with a resistive load Z_L identical to Z_0, the impedance 'seen' at the other end will also be Z_0. The condition for maximum power transfer is when the source impedance Z_S is also real and equivalent to both Z_0 and Z_L. Therefore, for maximum power transfer

$$Z_S = Z_0 = Z_L \tag{1.26}$$

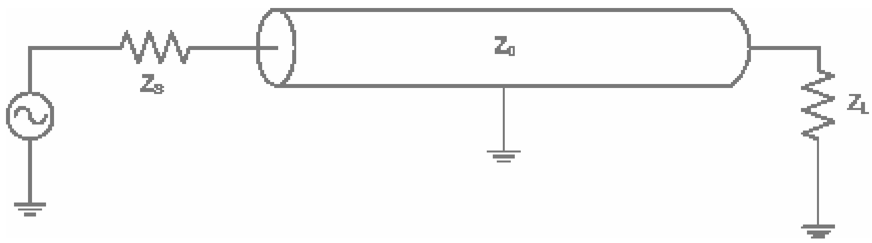

Figure 1-2 The fundamental transmission line configuration

This is shown schematically in Figure 1-2, in this case represented by a coaxial transmission line, though it would apply to any type of transmission line. The source is shown by a Thevenin equivalent circuit of source impedance Z_S.

The reciprocal of characteristic impedance is characteristic admittance Y_0, where:

$$Y_0 = \frac{1}{Z_0} \tag{1.27}$$

Although Z_0 is in more common usage, Y_0 is often useful when dealing with matching problems where the design requires switching between impedances and admittances.

1.7 Voltage Reflection Coefficient, Voltage Standing Wave Ratio and Return Loss

Most practical systems which include transmission lines will inevitably have some degree of mismatch. A mismatch occurs in a transmission line when either Z_S or Z_L differs from

Z_0. When a transmission line is not perfectly matched the mismatch may be measured by the voltage reflection coefficient, ρ_V. This is a quantity represented in magnitude and phase by the complex ratio of the reflected voltage wave (V_R) divided by the forward voltage wave (V_R), or:

$$\rho_V = \frac{V_R}{V_F} = \frac{Be^{\gamma z}}{Ae^{-\gamma z}} = Ce^{2\gamma z} \qquad (1.28)$$

where C is also a constant. A similar expression may be obtained for the current reflection coefficient, but this is less frequently used. It is normally understood that the simple expression 'reflection coefficient' refers in fact to the voltage reflection coefficient.

The more perfectly the transmission line is matched, the smaller the reflected wave will be. Therefore ρ_V will approach zero magnitude for a perfectly matched line. For the unmatched transmission line, the presence of z in the exponent indicates that the reflection coefficient varies with the position along the line at which it is measured. If the line is loss free ($\alpha = 0$ in (1.15)) the exponent becomes purely imaginary and the magnitude of the reflection coefficient $|\rho_V|$ is unchanged at any position along the line. However, its phase will vary along the line.

An unmatched transmission line would set up a standing wave: one in which V_F and V_R are finite. The standing wave is the resultant wave formed from the forward and reverse waves, V_F and V_R respectively. The standing wave variation with position on the line would be described by (1.13).

A common scalar measure of the degree of mismatch of a transmission line is the voltage standing wave ratio (VSWR), whose symbol is the lower case 's'. This is defined as the ratio of the voltage maxima to voltage minima, so

$$s = \frac{|V_F| + |V_R|}{|V_F| - |V_R|} = \frac{1 + V_R / V_F}{1 - V_F / V_R} = \frac{1 + |\rho_V|}{1 - |\rho_V|} \qquad (1.29)$$

VSWR is often expressed as a ratio in colon notation relative to unity. For example if $s = 1.15$, then $VSWR = 1.15 : 1$. (1.29) relates the VSWR to the *magnitude* of the voltage reflection coefficient. The VSWR is a linear scalar measure of match quality. It does not have any units and is always greater than or equal to unity which represents a perfect match. Alternatively the reflection coefficient magnitude may be expressed in terms of the VSWR by reorganising (1.29) as follows:

$$|\rho_V| = \frac{s-1}{s+1} \qquad (1.30)$$

A commonly used logarithmic scalar measurement of mismatch is known as return loss (RL) and is defined as:

$$RL = \left| 20 \log_{10} |\rho_V| \right| \quad dB \qquad (1.31)$$

Return loss is measured in decibels (dB) as a positive quantity in the same way as the loss of an attenuator is a positive quantity. The larger the return loss the better the match.

1.8 The Variation of Impedance Along a Practical (Lossy) Mismatched Transmission Line

For a practical (lossy) transmission line, mismatched at the distant end, the voltage reflection coefficient will vary at points along the line in both magnitude and phase.

If the length of the line is l, then substitution into (1.28) for $z=0$ and $z=l$ will yield the following equation for the reflection coefficient measured at the input to the line (ρ_0) in terms of the reflection coefficient at the load connected to the end of the line (ρ_T).

$$\rho_0 = \rho_T e^{-2\gamma l} \tag{1.32}$$

Dividing (1.19) by (1.23) gives the general expression relating the impedance Z_T connected to the end of the line with the forward and reflected waves V_F and V_R respectively.

$$\frac{V}{Z_0 I} = \frac{Z_T}{Z_0} = \frac{V_F + V_R}{V_F - V_R} = \frac{1+\frac{V_R}{V_F}}{1-\frac{V_R}{V_F}} = \frac{1+\rho_T}{1-\rho_T} \tag{1.33}$$

Notice that all terms in this equation are actual complex quantities. In terms of ρ_T (1.33) becomes:

$$\rho_T = \frac{Z_T - Z_0}{Z_T + Z_0} \tag{1.34}$$

Substituting for ρ_T into (1.32) gives

$$\rho_l = \left(\frac{Z_T - Z_0}{Z_T + Z_0}\right) e^{-2\gamma l} \tag{1.35}$$

where ρ_l is the reflection coefficient at the input of a line of length l terminated in an impedance Z_T. A similar equation to may be written but this time relating the input impedance of line Z_{IN} to the associated reflection coefficient ρ_0:

$$\frac{Z_{IN}}{Z_0} = \frac{1+\rho_0}{1-\rho_0} \tag{1.36}$$

Substituting for ρ_0 from (1.32) and expanding

$$\begin{aligned}\frac{Z_{IN}}{Z_0} &= \frac{1+\rho_T e^{-2\gamma l}}{1-\rho_T e^{-2\gamma l}} = \frac{1+\left(\frac{Z_T - Z_0}{Z_T + Z_0}\right)e^{-2\gamma l}}{1-\left(\frac{Z_T - Z_0}{Z_T + Z_0}\right)e^{-2\gamma l}} = \frac{Z_T + Z_0 + \left(Z_T - Z_0\right)e^{-2\gamma l}}{Z_T + Z_0 - \left(Z_T - Z_0\right)e^{-2\gamma l}} \\ &= \frac{Z_T + Z_0\left(\frac{1-e^{-2\gamma l}}{1+e^{-2\gamma l}}\right)}{Z_0 + Z_T\left(\frac{1-e^{-2\gamma l}}{1+e^{-2\gamma l}}\right)} = \frac{Z_T + Z_0 \tanh\gamma l}{Z_0 + Z_T \tanh\gamma l}\end{aligned} \tag{1.37}$$

The resulting equation is often known as the (lossy) transmission line equation:

$$Z_{IN} = Z_0 \left(\frac{Z_T + Z_0 \tanh \gamma l}{Z_0 + Z_T \tanh \gamma l} \right) \tag{1.38}$$

This is used to calculate the input impedance of a lossy transmission line of characteristic impedance Z_0, length l, propagation coefficient γ whilst it is terminated with an impedance Z_T.

1.9 The Loss-Free Transmission Line

1.9.1 The Loss-Free Transmission Line Equation

Often in transmission line problems it is adequate to assume the line to be loss free as this adds several very welcome simplifications in the algebra such as dealing with trigonometric functions instead of hyperbolic functions.

For a loss free transmission line the attenuation constant α will be zero, so (1.15) becomes

$$\gamma = \alpha + j\beta = j\beta \tag{1.39}$$

Substituting for $\gamma = j\beta$ in the loss-free case and using the Euler identities

$$\begin{aligned} e^{jx} + e^{-jx} &= 2\cos x \\ e^{jx} - e^{-jx} &= 2j\sin x \end{aligned} \tag{1.40}$$

the hyperbolic tangent simplifies as follows

$$\begin{aligned} \tanh \gamma l &= \frac{1-e^{-2\gamma l}}{1+e^{-2\gamma l}} = \frac{1-e^{-j2\beta l}}{1+e^{-j2\beta l}} \\ &= \frac{e^{j\beta l} - e^{-j\beta l}}{e^{j\beta l} + e^{-j\beta l}} = \frac{2j\sin\beta l}{2\cos\beta l} \\ &= j\tan\beta l \end{aligned} \tag{1.41}$$

Performing this substitution yields the transmission line equation for the loss-free case

$$Z_{IN} = Z_0 \left(\frac{Z_T + jZ_0 \tan \beta l}{Z_0 + jZ_T \tan \beta l} \right) \tag{1.42}$$

This may be used to calculate the impedance looking into a loss-free transmission line of characteristic impedance Z_0, length l and terminated at the distant end with an impedance of Z_T.

1.9.2 Phase Constant and Phase Velocity

In this case the phase constant β is given by

$$\beta = \frac{2\pi}{\lambda} \quad rad/m \tag{1.43}$$

where λ is the wavelength within the transmission line. This equation is used to determine the spatial phase variation (βl) along a transmission line.

For the loss free case $R = 0$ and $G = 0$ therefore, by equating (1.14) with (1.39):

$$j\beta = \sqrt{j^2\omega^2 LC}$$
$$\beta = \omega\sqrt{LC} \qquad (1.44)$$

Therefore, equating (1.43) with (1.44) and using the relationship relating temporal and angular frequency:

$$\omega = 2\pi f \qquad (1.45)$$

yields

$$\beta = \frac{2\pi}{\lambda} = \omega\sqrt{LC}$$
$$\frac{2\pi}{\lambda} = 2\pi f\sqrt{LC}$$
$$\frac{1}{\lambda f} = \sqrt{LC} \qquad (1.46)$$
$$v_p = \lambda f = \frac{1}{\sqrt{LC}}$$

where v_P is the phase velocity of propagation within the transmission line. If the transmission line is air spaced then v_P is identical to the velocity of electromagnetic radiation in free space ($v_p = c$). If it includes a solid dielectric of relative permittivity (dielectric constant) ε_r, then v_P is related to c by

$$v_p = \frac{c}{\sqrt{\varepsilon_r}} \qquad (1.47)$$

If the dielectric is partially air-spaced, then it may be assigned an effective dielectric constant k_{eff} where

$$v_p = \frac{c}{\sqrt{k}} \qquad (1.48)$$

The value of k would be found by measurement to be a value between unity (air only dielectric) and ε_r (solid dielectric).

The assumption in these definitions is that the phase velocity is constant and not a function of frequency.

2 WAVEGUIDES

A waveguide is a transmission line constructed to guide electromagnetic energy from one location to another. There are two forms of practical waveguide: dielectric-conductor and dielectric-dielectric. The dielectric-conductor waveguide comprises a conductor surrounding a dielectric through which the radio frequency (RF) energy is directed. A dielectric-dielectric waveguide is of similar construction with one of the dielectrics surrounding the other, and the two dielectrics have differing refractive indices, relative permittivities or dielectric constants.

2.1 The Rectangular Waveguide [2]

The rectangular waveguide is a form of very low loss transmission line used across the range of microwave frequencies from approximately typically from about 1 GHz to over 220 GHz. It comprises a rigid, precision rectangular hollow pipe constructed from a good electrical conductor with the inner hollow formed from a dielectric material. The dielectric must exhibit a suitably low loss at the maximum frequency to be propagated through the waveguide. Typically this might be dry air at normal atmospheric pressure or another gas such as dry nitrogen at a slightly elevated pressure compared to atmospheric. The walls are often constructed from copper, as this is a good electrical conductor, relatively cheap and chemically stable. Sometimes the internal walls may be plated with silver. Silver is a better electrical conductor that copper and, provided that the plating thickness is sufficient, the transmission loss will be slightly less that that with an equivalent copper waveguide. Although rectangular waveguide is expensive and bulky compared to the alternative transmission lines such as microstrip or coaxial cables, it is used widely for high power transmitter feeds, millimetre wave components and precision measuring equipment.

A difference between a waveguide and a TEM transmission line of the type described in Section 1.3 is that the waveguide comprises a single conductor so there is no possible path to form an electrical circuit requiring separate ‘go’ and ‘return’ current paths. It cannot therefore support TEM waves. Unlike TEM waves which propagate down to DC, a rectangular waveguide will only support frequencies above a threshold and then only in the form or transverse electric (TE) waves or transverse magnetic (TM) waves.

The geometry of a cross section of the rectangular waveguide is shown in Figure 2-1. The longer and the shorter internal transverse dimensions are a and b respectively ($a \succ b$). For the most common type of rectangular waveguide $a = 2b$. The x and y axes are aligned with the longer and shorter dimensions respectively. The direction of propagation through the waveguide, is parallel to the z axis and mutually perpendicular to both the x and y axes. The electrical properties of the material within the guide are described by its permittivity ε and permeability μ, where

$$\varepsilon = \varepsilon_0 \varepsilon_r \qquad (2.1)$$

and

$$\mu = \mu_0 \mu_r \qquad (2.2)$$

In (2.1) and (2.2)

μ_0 is the absolute permeability

μ_r is the relative permeability

ε_0 is the absolute permittivity

ε_r is the relative permittivity also known as the dielectric constant

In most cases, for a practical waveguides, the filling will be of a non-magnetic ($\mu_r = 1$).

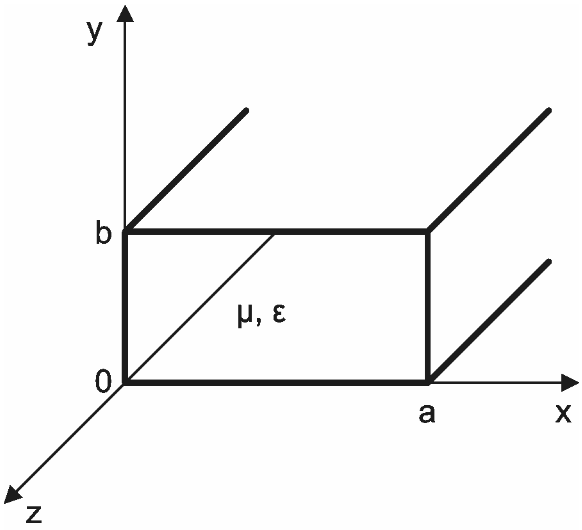

Figure 2-1 The rectangular co-ordinate system used with rectangular waveguides

The differential form of Faraday's Law, also known as the Maxwell - Faraday equation, is

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} = -\mu \frac{\partial \boldsymbol{H}}{\partial t} \tag{2.3}$$

in vector form, where the bold type indicates a vector quantity in this and all subsequent equations. $\boldsymbol{E}$ is the electric field vector and $\boldsymbol{H}$ is the magnetic field vector. The magnetic field is a time varying quantity so its temporal (time dependent) phase may be expressed, for example, in the following exponential form

$$H = H_0 e^{j\omega t} \tag{2.4}$$

where ω is the angular frequency in radians per second (rad/s), t is the time in seconds and H_0 is the value of H at $t = 0$.

Differentiating (2.4) with respect to t gives

$$\frac{\partial H}{\partial t} = j\omega H_0 e^{j\omega t} = j\omega H \tag{2.5}$$

Similarly, the electric field is also time varying, so

$$\frac{\partial E}{\partial t} = j\omega E_0 e^{j\omega t} = j\omega E \tag{2.6}$$

Therefore the Maxwell-Faraday equation becomes

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} = -\mu \frac{\partial \boldsymbol{H}}{\partial t} = -j\omega\mu \boldsymbol{H} \tag{2.7}$$

Another one of Maxwell's equations, also known as the Maxwell-Ampere equation, is

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{t}} = \boldsymbol{J} + \varepsilon \frac{\partial \boldsymbol{E}}{\partial t} \tag{2.8}$$

where $\boldsymbol{J}$ is the conduction current vector and $\boldsymbol{D}$ is the surface charge density.

For the dielectric in the hollow section of a rectangular waveguide, no conduction current can flow but an alternating current can, also known as a convection current. Therefore $\boldsymbol{J} = 0$ and (2.8) simplifies to

$$\nabla \times \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial \boldsymbol{t}} = \varepsilon \frac{\partial \boldsymbol{E}}{\partial t} = j\omega\varepsilon \boldsymbol{E} \qquad (2.9)$$

(2.7) and (2.9) are used to derive the differential equations that describe transverse components of both the electrical and magnetic fields as described in the next section.

2.1.1.1 Transverse Electric and Transverse Magnetic Fields

By expanding the curl expression in (2.7) and expressing the $\boldsymbol{H}$ field in terms of its rectangular vector components $\boldsymbol{H}_x$, $\boldsymbol{H}_y$ and $\boldsymbol{H}_z$:

$$\nabla \times \boldsymbol{E} = \begin{vmatrix} \boldsymbol{a_x} & \boldsymbol{a_y} & \boldsymbol{a_z} \\ \frac{\partial}{\partial \boldsymbol{x}} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} = -j\omega\mu\left(H_x\boldsymbol{a_x} + H_y\boldsymbol{a_y} + H_z\boldsymbol{a_z}\right) \qquad (2.10)$$

where $\boldsymbol{a}_x$, $\boldsymbol{a}_y$ and $\boldsymbol{a}_z$ are the unit vectors in the x, y and z directions respectively.

As propagation is in the z direction, the spatial phase dependence is also in the z direction, so a similar simplification that was adopted with the temporal part in (2.6) may be applied to the spatial part. An electric field E may be represented in the exponential spatial phase format as:

$$E = E_0 e^{-j\beta z} \qquad (2.11)$$

Differentiating this with respect to z gives

$$\frac{\partial E}{\partial z} = -j\beta E_0 e^{-j\beta z} = -j\beta E \qquad (2.12)$$

The vector curl equation in (2.10) may be expanded and substitutions made from (2.6) and (2.12) for the x, y and z coefficients of $\boldsymbol{E}$ and $\boldsymbol{H}$, followed by equating coefficients for each of the unit vectors to yield the following six equations.

$$\frac{\partial E_z}{\partial y} + j\beta E_y = -j\omega\mu H_x \qquad (2.13)$$

$$-j\beta E_x - \frac{\partial E_z}{\partial x} = -j\omega\mu H_y \qquad (2.14)$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu H_z \qquad (2.15)$$

$$\frac{\partial H_z}{\partial y} + j\beta H_y = j\omega\varepsilon E_x \qquad (2.16)$$

$$-j\beta H_x - \frac{\partial H_z}{\partial x} = j\omega\varepsilon E_y \qquad (2.17)$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\varepsilon E_z \qquad (2.18)$$

Equations (2.13) to (2.18) may be solved for the transverse field components of $\boldsymbol{E}$ (E_x and E_y) and $\boldsymbol{H}$ (H_x and H_y) to give the following

$$E_x = \frac{-j}{k_c^{\,2}}\left(\beta \frac{\partial E_z}{\partial x} + \omega\mu \frac{\partial H_z}{\partial y} \right) \tag{2.19}$$

$$E_y = \frac{j}{k_c^{\,2}}\left(-\beta \frac{\partial E_z}{\partial y} + \omega\mu \frac{\partial H_z}{\partial x} \right) \tag{2.20}$$

$$H_x = \frac{j}{k_c^{\,2}}\left(\omega\varepsilon \frac{\partial E_z}{\partial y} - \beta \frac{\partial H_z}{\partial x} \right) \tag{2.21}$$

$$H_y = \frac{-j}{k_c^{\,2}}\left(\omega\varepsilon \frac{\partial E_z}{\partial x} + \beta \frac{\partial H_z}{\partial y} \right) \tag{2.22}$$

where

$$k_c^{\,2} = k^2 - \beta^2 \tag{2.23}$$

k_c is the cutoff wavenumber or phase constant specific to the waveguide.

k is the phase constant of a plane (TEM) wave propagating in an unbounded medium electrically described by ε and μ.

β is the phase constant in the direction of propagation along the waveguide. Equation (2.23) may be expressed in terms of wavelengths by

$$\frac{1}{\lambda_c^{\,2}} = \frac{1}{\lambda^2} - \frac{1}{\lambda_g^{\,2}} \tag{2.24}$$

where

λ_c is the cutoff wavelength for the rectangular waveguide under consideration.

λ is the TEM wavelength for an equivalent unbounded plane wave considered as propagating through the medium described by ε and μ.

λ_g is the guide wavelength in the direction of propagation.

The results presented in (2.19) through (2.22) are the differential equations which define transverse electric (TE) fields and transverse magnetic (TM) fields.

2.1.1.2 Transverse Electric Waves

Transverse electric (TE) waves, by definition, are those for which $E_z = 0$ and $H_z \neq 0$ so the differential terms of E_z in (2.19) through (2.22) become zero, giving the following equations

$$E_x = \frac{-j\omega\mu}{k_c^{\,2}} \frac{\partial H_z}{\partial y} \tag{2.25}$$

$$E_y = \frac{j\omega\mu}{k_c^{\,2}} \frac{\partial H_z}{\partial x} \tag{2.26}$$

$$H_x = \frac{-j\beta}{k_c^2}\frac{\partial H_z}{\partial x} \quad (2.27)$$

$$H_y = \frac{-j\beta}{k_c^2}\frac{\partial H_z}{\partial y} \quad (2.28)$$

2.1.1.3 Transverse Magnetic Waves

Transverse magnetic (TM) waves, by definition, are those for which $E_z \neq 0$ and $H_z = 0$ so the differential terms of H_z in (2.19) through (2.22) become zero, giving the following equations

$$E_x = \frac{-j\beta}{k_c^2}\frac{\partial E_z}{\partial x} \quad (2.29)$$

$$E_y = \frac{-j\beta}{k_c^2}\frac{\partial E_z}{\partial y} \quad (2.30)$$

$$H_x = \frac{j\omega\varepsilon}{k_c^2}\frac{\partial E_z}{\partial y} \quad (2.31)$$

$$H_y = \frac{-j\omega\varepsilon}{k_c^2}\frac{\partial E_z}{\partial x} \quad (2.32)$$

2.1.1.4 TE and TM Modes

To examine the modes that may be supported by rectangular waveguides, we need to first return to the modified version of the Maxwell-Faraday equation (2.7).

Taking the curl of both sides of (2.7) and substituting the modified version of the Maxwell-Ampere equation from (2.9):

$$\nabla \times \nabla \times \boldsymbol{E} = -j\omega\mu\left(\nabla \times \boldsymbol{H}\right) = -j\omega\mu\left(j\omega\varepsilon \boldsymbol{E}\right) = \omega^2\mu\varepsilon\boldsymbol{E} \quad (2.33)$$

The next step is to apply the following vector identity for a general vector $\boldsymbol{A}$ to (2.33)**.**

$$\nabla \times \nabla \times \boldsymbol{A} = \nabla\left(\nabla \bullet \boldsymbol{A}\right) - \nabla^2\boldsymbol{A} \quad (2.34)$$

giving

$$\nabla\left(\nabla \bullet \boldsymbol{E}\right) - \nabla^2\boldsymbol{E} = \omega^2\mu\varepsilon\boldsymbol{E} \quad (2.35)$$

Since there is no stored charge, the volumetric charge density becomes zero ($\rho = 0$) and the Maxwell Gauss equation $\nabla \bullet \boldsymbol{D} = \rho$ also becomes zero, so

$$\varepsilon\nabla \bullet \boldsymbol{E} = 0 \quad (2.36)$$

and

$$\nabla \bullet \boldsymbol{E} = 0 \quad (2.37)$$

Substituting (2.37) into (2.35) gives the Helmholtz wave equation in terms of electric field:

$$\nabla^2\boldsymbol{E} + \omega^2\mu\varepsilon\boldsymbol{E} = 0 \quad (2.38)$$

A similar Helmholtz expression in terms of the magnetic field $\boldsymbol{H}$ can be obtained starting with the Maxwell-Ampere equation (2.8) applied to the internal medium of the waveguide which is an insulator so $\boldsymbol{J} = 0$ and the modified version (2.9) is used. A further substitution from (2.7) produces the result

$$\nabla^2 \boldsymbol{H} + \omega^2 \mu\varepsilon \boldsymbol{H} = 0 \qquad (2.39)$$

(2.39) may be reduced to its rectangular components since, for the left hand side

$$\nabla^2 \boldsymbol{H} \equiv \nabla^2 H_x \boldsymbol{a_x} + \nabla^2 H_y \boldsymbol{a_y} + \nabla^2 H_z \boldsymbol{a_z} \qquad (2.40)$$

and

$$\boldsymbol{H} \equiv H_x \boldsymbol{a_x} + H_y \boldsymbol{a_y} + H_z \boldsymbol{a_z} \qquad (2.41)$$

To apply the equations for the TE waves (2.25) to (2.28), the H_z component of the Helmholtz equation must be extracted from (2.40) and (2.41):

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + k^2 \right) H_z = 0 \qquad (2.42)$$

The spatial phase dependence of H_z in the z direction may be described by

$$H_z\left(x, y, z\right) = h_z(x, y)e^{-j\beta z} \qquad (2.43)$$

Since h_z is not actually a function of z, it may be treated as a constant when differentiating H_z with respect to z, so

$$\frac{\partial^2 H_z}{\partial z^2} = -\beta H_z \qquad (2.44)$$

Substituting (2.44) into (2.42) and also using $k_c^{\,2} = k^2 - \beta^2$ (2.23) gives the result

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_c^{\,2} \right) h_z\left(x, y\right) = 0 \qquad (2.45)$$

The partial differential equation (2.45) may be solved by the method known as ‘separation of the variables’, letting

$$h_z\left(x, y\right) = X\left(x\right)Y\left(y\right) \qquad (2.46)$$

then

$$\frac{\partial^2 h_z}{\partial x^2} = Y\frac{\partial^2 X}{\partial x^2} \qquad (2.47)$$

$$\frac{\partial^2 h_z}{\partial y^2} = X\frac{\partial^2 Y}{\partial y^2} \qquad (2.48)$$

Substituting (2.46), (2.47) and (2.48) into (2.45) gives the following result

$$\frac{1}{X}\frac{d^2 X}{dx^2} + \frac{1}{Y}\frac{d^2 Y}{dy^2} + k_c^{\,2} = 0 \qquad (2.49)$$

The principle of separation of the variables requires each of the terms in (2.49) to be equal to a constant thus generating two more differential equations

$$\frac{d^2 X}{dx^2} + k_x^{\,2} X = 0 \qquad (2.50)$$

$$\frac{d^2Y}{dy^2} + k_y{}^2 Y = 0 \qquad (2.51)$$

where k_x and k_y are constants such that

$$k_x{}^2 + k_y{}^2 = k_c{}^2 \qquad (2.52)$$

The solutions of (2.50) and (2.51) are

$$X = A\cos k_x x + B\sin k_x x \qquad (2.53)$$

and

$$Y = C\cos k_y y + D\sin k_y y \qquad (2.54)$$

where A, B, C and D are constants, so

$$h_z(x, y) = XY = (A\cos k_x x + B\sin k_x x)(C\cos k_y y + D\sin k_y y) \qquad (2.55)$$

2.1.1.5 Boundary Conditions for TE Modes

The waveguide walls are very good conductors so will not support any tangential components of electric field. Using the rectangular coordinate system given in Figure 2-1, this must occur at the waveguide walls as follows:

$$e_x(x, y) = 0 \quad at \quad y = 0 \quad and \quad y = b \qquad (2.56)$$

$$e_y(x, y) = 0 \quad at \quad x = 0 \quad and \quad x = b \qquad (2.57)$$

e_x and e_y are obtained by differentiating (2.55) with respect to both x and y and using results in (2.26) and (2.25) respectively.

$$\frac{\partial h_z}{\partial x} = k_x(-A\sin k_x x + B\cos k_x x)(C\cos k_y y + D\cos k_y y) \qquad (2.58)$$

$$\frac{\partial h_z}{\partial y} = k_y(A\cos k_x x + B\sin k_x x)(-C\sin k_y y + D\cos k_y y) \qquad (2.59)$$

By substituting (2.58) and (2.59) into (2.26) and (2.25) respectively, the solutions for e_x and e_y are

$$e_x = \frac{-j\omega\mu}{k_c{}^2} k_y(A\cos k_x x + B\sin k_x x)(-C\sin k_y y + D\cos k_y y) \qquad (2.60)$$

$$e_y = \frac{-j\omega\mu}{k_c{}^2} k_x(-A\sin k_x x + B\cos k_x x)(C\cos k_y y + D\cos k_y y) \qquad (2.61)$$

Using the boundary conditions in (2.56) and (2.57), from (2.60), $D = 0$ and

$$k_y = \frac{n\pi}{b} \quad for \quad n = 0, 1, 2... \qquad (2.62)$$

Similarly, from (2.61), $B = 0$ and

$$k_x = \frac{m\pi}{a} \quad for \quad m = 0, 1, 2... \qquad (2.63)$$

Using (2.43) the solution for H_z is therefore

$$H_z(x,y,z) = A_{mn}\cos\frac{m\pi x}{a}\cos\frac{n\pi y}{b}e^{-i\beta z} \qquad (2.64)$$

where A_{mn} is a constant originating from the constants A and C.

The solutions for the transverse E and H components are obtained by differentiating (2.64) with respect to x or y as appropriate and substituting into (2.25), (2.26), (2.27) and (2.28) to give the following equations.

$$E_x = \frac{j\omega\mu n\pi}{k_c^2 b}A_{mn}\cos\frac{m\pi x}{a}\sin\frac{n\pi y}{b}e^{-i\beta z} \qquad (2.65)$$

$$E_y = \frac{-j\omega\mu m\pi}{k_c^2 a}A_{mn}\sin\frac{m\pi x}{a}\cos\frac{n\pi y}{b}e^{-i\beta z} \qquad (2.66)$$

$$H_x = \frac{j\beta m\pi}{k_c^2 a}A_{mn}\sin\frac{m\pi x}{a}\cos\frac{n\pi y}{b}e^{-i\beta z} \qquad (2.67)$$

$$H_y = \frac{j\beta n\pi}{k_c^2 b}A_{mn}\cos\frac{m\pi x}{a}\sin\frac{n\pi y}{b}e^{-i\beta z} \qquad (2.68)$$

These equations will describe the TE_{mn} mode.

The propagation constant, from (2.23) by substituting (2.62) and (2.63) is

$$\beta = \sqrt{k^2 - k_c^2} = \sqrt{k^2 - \left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \qquad (2.69)$$

The propagation constant is real when

$$k \succ k_c \qquad (2.70)$$

and the cutoff condition is

$$k_c = \sqrt{\left(\frac{m\pi}{a}\right)^2 - \left(\frac{n\pi}{b}\right)^2} \qquad (2.71)$$

The general expression for the cutoff frequency f_{cmn} for any combination of m and n is

$$f_{cmn} = \frac{k_c}{2\pi\sqrt{\mu\varepsilon}} = \frac{1}{2\pi\sqrt{\mu\varepsilon}}\sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \qquad (2.72)$$

The fundamental mode is that with $m=1, \quad n=1$, known as TE_{10}, so

$$f_{c10} = \frac{1}{2a\sqrt{\mu\varepsilon}} \qquad (2.73)$$

3 REFERENCES

1. Chipman, Robert A.; Transmission Lines; Schaum's Outline Series in Engineering; McGraw Hill Book Company;
2. Pozar, David M.; *Microwave Engineering - Third Edition*; John Wiley & Sons Inc; ISBN 0-471-44878-8 (2005); pp 106 - 117.