THE QUEEN'S AWARDS
FOR ENTERPRISE
2001

BBC

R&D White Paper

WHP 126

September 2005

Co-axial cables

C. Gandy

Research & Development
BRITISH BROADCASTING CORPORATION

BBC Research & Development
White Paper WHP 126

Co-axial cables

Chris Gandy

Spectrum Planning Group

Abstract

The physical basis for the operation of co-axial transmission lines is ancient, well-understood and fundamental to radio-frequency electronics. However, some aspects of the link between the underlying physics and the practical implementation appear rather obscure; for example, the reasons for the widespread use of two particular characteristic impedances, 50 Ω and 75 Ω.

Also, interest is gathering in the topic of transmitter power ratios for switch-over from analogue to digital terrestrial television, and the capacity of the co-axial 'feeders' to handle the requisite average and peak powers features amongst a long list of critical factors that will affect the cost and practicality of switch-over.

This White Paper provides some of the necessary background about the electrical characteristics of co-axial cables. It is demonstrated that there are optimum characteristic impedances and they are evaluated, and it is shown that practical co-axial cables exhibit a fundamental peak-to-mean ratio; they can withstand short-term peaks of voltage considerably greater than the voltage corresponding to the long-term maximum power rating.

Additional key words: co-axial cable, transmission line, characteristic impedance, power rating, dielectric breakdown, dielectric hysteresis, loss

White Papers are distributed freely on request.

Authorisation of the Chief Scientist is required for publication.

Co-axial cables

Chris Gandy

Spectrum Planning Group

1. Introduction

The work of Spectrum Planning Group is based on principles of antennas and propagation that are, for the most part, straightforward, well documented and well understood. Occasionally, however, questions arise for which the quantitative answer may be obtainable from available text books but, for one reason or another, the full explanation is obscure. Such appears to be the case in respect of the question: why do co-axial cables usually have characteristic impedances of 50 Ω or 75 Ω?

The characteristic impedances of transmission lines is a matter so fundamental to RF electronics that one might expect it to be common knowledge, and well understood, why for the past 75 years or so most co-axial cables have been manufactured with impedances of 50 Ω or 75 Ω. But this appears not to be the case, and whilst factual statements of optimum impedances can be found in the literature [1], the full story of why those two particular values are ubiquitous seems more than a little obscure. It only adds to the mystery when the author of a well-respected text book on electromagnetics [2] sends the reader off to look up an article [3] in a supplement to a magazine from 1990 – maybe not the easiest of references to obtain – whilst a more-recent magazine article on the subject [4] is interesting but goes no further than 'scratching the surface'.

Whilst attempting to provide a clear answer to a colleague's question about the terminal impedances of antennas ("why do they have those particular values?") I ventured into some of the fundamentals of radiation resistance, but I also kept coming back to the fact that the world of RF electronics seems to revolve around those two standard impedances for co-axial cables: 50 Ω and 75 Ω. Surely, this deserved a clear, complete explanation if, indeed, that were possible.

Additionally, interest is gathering in the topic of power ratios for television switch over. Whilst it had been thought that each analogue transmission would be replaced by a DTT transmission at a relative power of -10 dB, or perhaps -7 dB (DTT mean power relative to PAL-VSB peak-sync power, per transmission), in some cases now the idea of -4 dB is being mooted in an effort to overcome difficulties in international co-ordination. Even at -7 dB, concerns have been expressed about the capacity of existing co-axial feeders and other components, especially in the case of the 200 or so stations that will transmit 3 commercial DTT multiplexes as well as the 3 public-service ones where presently they carry the 4 national analogue services. This White Paper may assist discussions on this topic by providing some of the background about power ratings of co-axial cables.

Although none of what I will present is original, I aim to 'dust off' the theory and add some value by delving a little deeper than the authors of the well-known text books could justify, perhaps even in some dark corners where the sun rarely shines!

2. Electric field strength

A thin straight wire of infinite length carrying a charge per unit length ρ_L is surrounded by a radial electric field of strength E_r at a radius r given by [5]:

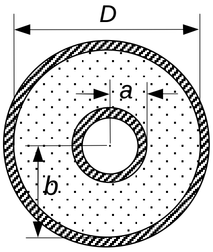
$$E_r = \frac{\rho_L}{2\pi\varepsilon r}$$

… where ε is the permittivity of the dielectric medium surrounding the wire. Generally $\varepsilon = \varepsilon_0 \varepsilon_r$ where ε_0 is the permittivity of space (8.85 pFm^{-1}) and ε_r is the dimensionless relative permittivity of the medium (also known as the dielectric constant); $\varepsilon_r = 1.0006$ for air which I'll take to be unity. This expression is derived from Coulomb's Law, which was originally found by experiment.

Because the field strength decays radially, there is a potential difference between any two points at different distances a and b from the wire:

$$\Delta V = V_a - V_b = \frac{\rho_L}{2\pi\varepsilon}\int_a^b \frac{\mathrm{d}r}{r} = \frac{\rho_L}{2\pi\varepsilon}\ln\frac{b}{a}$$

But there is no potential difference between any two points the same distance away so *equipotential* surfaces surround the wire in the form of co-axial cylinders. A cylindrical conductor can be inserted along an equipotential surface without necessarily distorting the electric field (it might just reduce its strength to zero on one side!), and if the electric field is developed between a co-axial pair of cylinders the result has the form of a co-axial transmission line.



If the radius of the outer surface of the inner conductor is a, the radius of the inner surface of the outer conductor is b, and a voltage V (shorthand for a potential difference, ΔV) is applied between the conductors:

$$V = \frac{\rho_L}{2\pi\varepsilon}\ln\frac{b}{a} \quad \text{so} \quad \rho_L = \frac{2V\pi\varepsilon}{\ln b/a} \quad \text{and then} \quad E_r = \frac{V}{r\ln b/a}$$

The field strength is greatest where r has its smallest value, immediately adjacent to the outer surface of the inner conductor where $r = a$ and:

$$E_a = \frac{V}{a\ln b/a}$$

When dealing with co-axial cables, two important parameters are the ratio of the radii, b/a for which I'll use the symbol ψ hereafter, and the diameter of the outer conducting surface which I'll denote as $D = 2b$, so $a = D/2\psi$ and:

$$\frac{V}{E_a} = \frac{D}{2}\cdot\frac{\ln\psi}{\psi}$$

If D is held constant and ψ is varied over the most-practical range $1.05 \le \psi \le 10$, the value of this expression will vary from $0.023D$ to $0.12D$ via a range of larger numbers such as $0.16D$ at $\psi = 5$ so it must pass through a maximum. At this point, the differential of the expression with respect to ψ will be zero.

Using $\frac{\mathrm{d}}{\mathrm{d}x}uv = u\frac{\mathrm{d}v}{\mathrm{d}x} + v\frac{\mathrm{d}u}{\mathrm{d}x}$, $\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{u}\right) = -\frac{1}{u^2}$ and $\frac{\mathrm{d}}{\mathrm{d}x}\ln u = \frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}x}$, and treating D as a constant:

$$\frac{\mathrm{d}}{\mathrm{d}\psi}\left(\frac{V}{E_a}\right) = \frac{D}{2}\left(-\frac{\ln\psi}{\psi^2} + \frac{1}{\psi^2}\right) = \frac{D}{2\psi^2}(1-\ln\psi) \quad \ldots \text{(deep sigh!)}$$

The differential is equal to zero when $\ln\psi = 1$ so $\psi = e = 2.718$. Then for a given maximum strength of the electric field in the co-axial line E_a, for example that which would lead to breakdown of the dielectric between the conductors, the greatest voltage can be tolerated when the ratio of the radii or diameters of the conducting surfaces is equal to e. With that condition, $V = E_a D/2e$ so the absolute value of the maximum voltage is proportional to the diameter of the line, or $E_a = 2Ve/D$ so the absolute field strength is reduced when the diameter of the transmission line is increased, both of which are fairly obvious.

Of course, this treatment has considered the electrostatic case but there is no reason for the field configuration to be appreciably different if an alternating voltage were applied, as long as all radial dimensions are kept much smaller than the wavelength, so the result should be equally applicable to the (AC) transmission-line case.

3. Transmission lines

However, the big difference between the two cases is in the dynamics of the latter: the flow of power through a transmission line, along with the attendant current, characteristic impedance (ratio of voltage to current), attenuation, velocity and dispersion. The fundamental theory behind the operation of transmission lines is covered to different extents in many of the well-known text books (e.g. [6]) so I will recap it only briefly here.

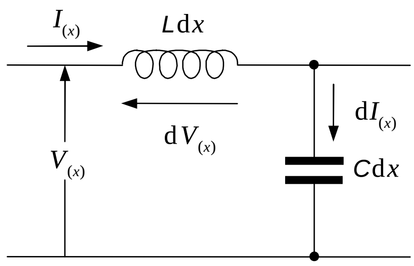
A transmission line is used to connect an AC generator to a load and it is an observed fact that propagation of power through a transmission line, as through any other medium, has a finite velocity so it is subjected to a time delay, from one end of the line to the other. In the simplest case of a generator producing a voltage that alternates sinusoidally, the variation of voltage in the line with distance away from the generator (or delay time) follows a sinusoidal wave that travels along the line in the direction of power flow. I will illustrate this in a moment.

Equally, the variation of current in the line can be represented by a travelling wave, as can the instantaneous power (product of voltage and current). These waves travel at a velocity v (metre s^{-1}) whilst the alternating frequency is ω (radian s^{-1}) so the wavelength in the line is given by $\lambda = 2\pi\, v/\omega$ (metre). Any more-complicated waveform can be decomposed into sinusoids of different frequencies.

It is useful to define $\beta = \omega / v = 2\pi/\lambda$ (radian metre^{-1}), the so-called 'propagation phase constant', sometimes referred to as the 'wave number' and given the symbol 'k'. This signifies the amount of phase excursion with distance down the line giving the phase retardation βx (relative to the phase of the generator) at any chosen point a distance x from the end to which the generator is connected.

The simplest form of transmission line consists of two conductors separated from one another so the line exhibits series inductance and shunt capacitance. The simplest case to analyse is a lossless uniform line of infinite length; then the inductance and capacitance can be quantified as *distributed* constants, that is L (Hm^{-1}) and C (Fm^{-1}) fundamentally, although smaller units would apply in practice such as nHm^{-1} and pFm^{-1}, as will be seen. Since the line is uniform along its length, there's no reason for the dynamics of power flow to be different from one point to another so the length of the line can be subdivided into many small sections and the analysis of one section should be applicable to the whole line by considering a cascade of such small sections. Taken to the limit, where each section has infinitesimal length (or, at least, length much smaller than λ), the section can be analysed using lumped circuit elements.

Considering an infinitesimal length of the line, dx, a distance x (metre) from the generator end, possessing an inductance $L\,\mathrm{d}x$ and a capacitance $C\,\mathrm{d}x$, and arranging these as a diagonally-inverted 'L' circuit excited by a sinusoidal wave of voltage $V_{(x)}$:



- the current $I_{(x)}$ passing through the inductance gives rise to a voltage drop $\mathrm{d}V_{(x)}$ between the input and output terminals of this section of line, on account of the inductive reactance, where $\mathrm{d}V_{(x)} = \mathrm{j}\omega L I_{(x)}\mathrm{d}x$;
- and the voltage developed across the capacitance ($V_{(x)}$ - $\mathrm{d}V_{(x)}$) causes a current to flow through it so, compared to the input current, the output current is reduced by $\mathrm{d}I_{(x)} = \mathrm{j}\omega C\mathrm{d}x\,(V_{(x)} - \mathrm{d}V_{(x)}) = \mathrm{j}\omega C V_{(x)}\mathrm{d}x$ neglecting the vanishingly-small term containing $\mathrm{d}x\mathrm{d}V_{(x)}$.

So: $$\frac{\mathrm{d}V_{(x)}}{\mathrm{d}x} = \mathrm{j}\omega L\, I_{(x)} \quad \text{and} \quad \frac{\mathrm{d}I_{(x)}}{\mathrm{d}x} = \mathrm{j}\omega C\, V_{(x)}$$

These are sometimes referred to as the 'telegraph equations' or the 'telegraphers' equations' though I don't suppose the average 19th-century telegrapher cared much about the theory. The right-hand sides are sometimes given with minus signs but I've used the convention normally applied when using Ohm's law in a circuit: $V = IR$ with the direction of positive V opposite that of positive I everywhere other than at the generator terminals.

Differentiating each of these, in turn, and inserting the other:

$$\frac{\mathrm{d}^2V_{(x)}}{\mathrm{d}x^2} = (\mathrm{j}\omega L)\frac{\mathrm{d}I_{(x)}}{\mathrm{d}x} = (\mathrm{j}\omega L)\mathrm{j}\omega C V_{(x)} \;\text{ so }\; \frac{\mathrm{d}^2V_{(x)}}{\mathrm{d}x^2} - (\mathrm{j}\omega L)(\mathrm{j}\omega C)V_{(x)} = 0 \;\text{ or }\; \frac{\mathrm{d}^2V_{(x)}}{\mathrm{d}x^2} + \omega^2 LCV_{(x)} = 0$$

$$\frac{\mathrm{d}^2I_{(x)}}{\mathrm{d}x^2} = (\mathrm{j}\omega C)\frac{\mathrm{d}V_{(x)}}{\mathrm{d}x} = (\mathrm{j}\omega C)\mathrm{j}\omega L I_{(x)} \;\text{ so }\; \frac{\mathrm{d}^2I_{(x)}}{\mathrm{d}x^2} - (\mathrm{j}\omega C)(\mathrm{j}\omega L)I_{(x)} = 0 \;\text{ or }\; \frac{\mathrm{d}^2I_{(x)}}{\mathrm{d}x^2} + \omega^2 LCI_{(x)} = 0$$

… and these are known as the transmission-line wave equations for $V_{(x)}$ and $I_{(x)}$.

A solution to the wave equation for $V_{(x)}$ is found in the form $V = e^{\gamma x}$ because then:

$$\frac{\mathrm{d}^2V}{\mathrm{d}x^2} = \gamma^2 e^{\gamma x} = \gamma^2 V \quad \text{and} \quad \left(\gamma^2 + \omega^2 LC\right)V = 0 \quad \text{so} \quad \gamma^2 + \omega^2 LC = 0$$

The roots of this are $\gamma = \pm\mathrm{j}\omega\sqrt{LC}$ so $V_{(x)}$ can be written as $V_{(x)} = \mathrm{K}_1 e^{\mathrm{j}\omega\sqrt{LC}x} + \mathrm{K}_2 e^{-\mathrm{j}\omega\sqrt{LC}x}$

… where K_1 and K_2 are constants with respect to x, although they may vary with respect to time.

$V_{(0)}$ = $K_1 + K_2$ so together they represent the instantaneous voltage on the line at $x = 0$ and they can be written as sinusoids[*] with respect to time: $K_1 = V_1 e^{j\omega t}$ and $K_2 = V_2 e^{j\omega t}$ where V_1 and V_2 are constants.

γ is known as the 'propagation constant' which, in general, is complex having real and imaginary parts: $\gamma = \alpha + j\beta$, where α is the 'attenuation constant' (a voltage ratio expressed in neper m^{-1}, which I'll explain later) and β the 'phase constant' (radian m^{-1}), as before. For the present, lossless, case $\alpha = 0$ and $\beta = \omega\sqrt{LC}$ so, inserting β and substituting for K_1 and K_2 we obtain:

$$V_{(x)} = V_1 e^{j\omega t} e^{j\beta x} + V_2 e^{j\omega t} e^{-j\beta x} = V_1 e^{j(\omega t + \beta x)} + V_2 e^{j(\omega t - \beta x)}$$

The second term $V_2 e^{j(\omega t - \beta x)}$ represents a wave for which passage along the line away from the generator (increasing x) tends to counteract the passage of time (increasing t) so it is a wave travelling in the forward, positive x direction; generator to load. Dividing through by β, the term can be written $V_2 e^{j\beta\{(\omega/\beta)t - x\}}$. Then if the exponent is set equal to some constant, $(\omega/\beta)t - x = K$, the differential of this equation with respect to time is $\omega/\beta - dx/dt = 0$ so $dx/dt = \omega/\beta = v$, the velocity of the travelling wave as noted earlier.

The first term $V_1 e^{j(\omega t + \beta x)}$ represents a wave travelling in the opposite direction, from the load back to the generator. This is said to carry power reflected by the load on account of any mismatch between its impedance and the characteristic impedance of the line. In the absence of any anisotropic material, the transmission line must physically allow waves to travel in both directions, and mathematically the possibility of this only arises because of the sign ambiguity of the square root. But I still find it remarkable that simple theory like this, with a token effort at maintaining generality, predicts something rather beyond what we modelled by placing a generator at one end. For the remainder of this document I will concentrate on the forward wave.

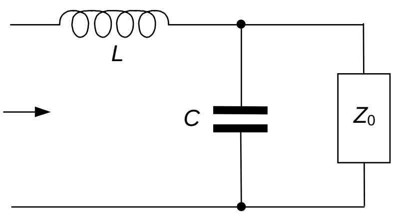
A solution of the same form can be found for the current wave equation.

4. Characteristic impedance

The ratio $V_{(x)}/I_{(x)}$ for a single wave is known as the 'characteristic impedance' of the line:

$$Z_0 = \frac{V_{(x)}}{I_{(x)}} = \frac{\dfrac{dI_{(x)}}{dx}\Big/-j\omega C}{\dfrac{dV_{(x)}}{dx}\Big/-j\omega L} = \frac{dI_{(x)}}{dV_{(x)}} \cdot \frac{L}{C} \quad \text{and if this, } Z_0 \text{ is a simple constant then } \frac{dI_{(x)}}{dV_{(x)}} = \frac{1}{Z_0} \text{ so } Z_0^{\,2} = \frac{L}{C}$$

… and $Z_0 = \sqrt{\dfrac{L}{C}}$.



The principle of cascading any number of small sections of uniform line can now be demonstrated by calculating the impedance Z_i presented by a single inverted-'L' section, representing an infinitesimal section of the line, terminated in an impedance of Z_0, as illustrated to the right. To simplify the working I have written L and C instead of $L\,dx$ and $C\,dx$:

$$Z_i = j\omega L + \frac{Z_0(1/j\omega C)}{Z_0 + 1/j\omega C} = j\omega L + \frac{Z_0}{1 + j\omega Z_0 C} = \frac{Z_0(1-\omega^2 LC) + j\omega L}{1 + j\omega Z_0 C}$$

$$= \frac{Z_0(1-\omega^2 LC) + j\omega L}{1 + j\omega Z_0 C} \cdot \frac{1 - j\omega Z_0 C}{1 - j\omega Z_0 C} = \frac{Z_0(1-\omega^2 LC) + \omega^2 Z_0 LC + j\omega L - j\omega Z_0^{\,2} C(1-\omega^2 LC)}{1 + \omega^2 Z_0^{\,2} C^2}$$

… but L and C are very small so their product, or the square of one of them, would be vanishingly small (remember, they're really $L\,dx$ and $C\,dx$) and such terms can be omitted, leaving:

$$Z_i = Z_0 + j\omega L - j\omega Z_0^{\,2} C \quad \text{but} \quad Z_0^{\,2} = L/C \quad \text{so} \quad Z_i = Z_0 + j\omega L - j\omega L$$

… and $Z_i = Z_0$. QED.

[*] The exponential form $e^{j\omega t} = \cos\omega t + j\sin\omega t$ is more convenient here than the trigonometric sin or cos. Either the real or imaginary part of $e^{j\omega t}$ can be used to represent a voltage varying sinusoidally with respect to time, and the phase discrepancy is unimportant in this case, but it probably helps to treat K_1 and K_2 consistently.

Of course, a lumped-element inverted-‘L’ section made from components having finite, rather than infinitesimal, values would behave as a low-pass filter with a cut-off frequency, and a resonance, at $\omega = 1/\sqrt{LC}$. In practice, losses would prevent the resonance from having an infinite effect, that is, the Q-factor would be limited. But for a uniform transmission line, having uniformly distributed inductance and capacitance, no part of the line has greater significance than any other and the cut-off frequency of the cascade of infinitesimal sections is at infinite frequency. There is, however, another cut-off frequency that matters, related to waveguide propagation, which I will describe later.

Capacitance is the ratio of charge to voltage, $C = Q / V$ and the capacitance per unit length of a co-axial line $C = \rho_L / V$, where ρ_L is the charge per unit length as before, so:

$$C = \frac{2\pi\varepsilon}{\ln\psi}$$

The self-inductance L of some system of conductors connected in a circuit is defined as the ratio of the total magnetic flux linkage (linking with these conductors) to the current I flowing through them. The flux density B_r at a radius r from a long straight conductor is given by [7]:

$$B_r = \frac{\mu I}{2\pi r}$$

… where μ is the permeability of the medium surrounding the conductor. Generally, $\mu = \mu_0\mu_r$ where μ_0 is the permeability of space (400π or 1260 nHm^{-1}, by definition) and μ_r is the dimensionless relative permeability of the medium. This expression is derived from the Biot-Savart law, also initially found by experiment, and indicates constant flux density at a given radius with no dependence on the radius of the conductor (other than that it must be less than r, as will be seen in a moment).

Considering a circular path of some radius R centred on the conductor and an infinitesimal element of length $\mathrm{d}\ell$ around this path, the integral of B around this path, that is the *line integral* of B (signified by a circle on the integral symbol) is given by:

$$\oint B.\mathrm{d}\ell = B\oint \mathrm{d}\ell = \frac{\mu I}{2\pi R}\oint \mathrm{d}\ell$$

The total path length is $2\pi R$ so: $$\oint B.\mathrm{d}\ell = \mu I$$

This is a variation on Ampère’s circuital law which equates the line integral of the magnetic field for any shape of closed path to the current I enclosed. More generally, $\mathrm{d}\ell$ would need to be represented by a vector because it has direction as well as (infinitesimal) magnitude, but in this case the circular symmetry and the uniformity of B around the integration path allows the use of scalars.

Applying this to a uniform hollow cylindrical conductor, if the radius of the path of integration is made smaller than the radius of the inner surface of this conductor then no current is enclosed so the line integral of the magnetic field must be zero in the space inside the conductor. This means either that the magnetic field must have zero magnitude or that it is arranged with positive and negative cycles around the path that cancel in the integration. For the circularly symmetric case in hand, the former is far more likely!

Thus the magnetic field, or flux density, between the conductors of a co-axial transmission line is approximately the same as would be found surrounding the inner conductor alone. The approximation is on account of the magnetic properties of the outer conductor but commonly-used materials like copper and aluminium have no substantial magnetic properties and, consequently, values of μ_r very close to unity.

The total flux linkage Λ, per unit length of line, is given by the integral of B_r over $a \le r \le b$:

$$\Lambda = \int_a^b B_r\,\mathrm{d}r = \frac{\mu I}{2\pi}\int_a^b \frac{\mathrm{d}r}{r} = \frac{\mu I}{2\pi}\ln\frac{b}{a}$$

The classical interpretation is the 'number of lines of force … outside the wire which will collapse upon the wire when the current ceases' [8]. Actually, 'lines of force' can be considered a misnomer because their arrangement is rings around the conductor and the force between two such conductors would be directed perpendicularly to these rings. Perhaps contours of equal force is a better description, like equipotentials in the electrostatic case.

Then the inductance per unit length of the line is:

$$L = \frac{\Lambda}{I} = \frac{\mu}{2\pi}\ln\psi \quad \text{or} \quad L = \frac{\ln\psi}{2\pi/\mu}$$

Substituting the expressions for L and C in a co-axial line into that for Z_0 for a simple, lossless, line:

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{1}{2\pi}\sqrt{\frac{\mu}{\varepsilon}}\ln\psi$$

Inserting values: $\mu = \mu_0\mu_r$ with $\mu_r = 1$, typical of most materials that would be used as dielectric separators in co-axial lines, and $\varepsilon = \varepsilon_0\varepsilon_r$ with $\varepsilon_0 = 8.85$ pFm^{-1}:

$$Z_0 = \frac{60}{\sqrt{\varepsilon_r}}\ln\psi$$

It then follows that $Z_0 = 60/\sqrt{\varepsilon_r}$ ohms to minimise the electric field strength in a co-axial line so, for a given dielectric strength, this characteristic impedance will allow the maximum alternating voltage between the conductors.

5. Maximum power handling on account of dielectric breakdown

However, the power P flowing through a transmission line is related to the peak (sinusoidal) voltage V across it by $P = V^2/2Z_0$ and the presence of Z_0 in the denominator means maximum power will coincide with a different value of Z_0 (in the same way that the peak torque and peak power of an engine occur at different RPM because power = torque × RPM).

$$V = E_a\frac{D}{2}\cdot\frac{\ln\psi}{\psi} \quad \text{so} \quad P = \frac{1}{2Z_0}\left(E_a\frac{D}{2}\cdot\frac{\ln\psi}{\psi}\right)^2 = \frac{(E_a D\ln\psi)^2}{8\psi^2 Z_0}$$

Inserting the expression for Z_0 in a lossless line:

$$P = \frac{\sqrt{\varepsilon_r}\,(E_a D\ln\psi)^2}{480\psi^2\ln\psi} = \frac{E_a{}^2\sqrt{\varepsilon_r}\,D^2}{480}\cdot\frac{\ln\psi}{\psi^2}$$

So, for a tolerable maximum field strength E_a, the maximum power is dependent on this different function of ψ. If D is held constant and ψ varied over the range $1.05 \le \psi \le 10$ again, the value of the variable part of this function will vary from 0.04 to 0.02 via larger numbers such as 0.06 at $\psi = 5$ so it must pass through a maximum which, as before, can be found by differentiating.

$$\frac{\mathrm{d}}{\mathrm{d}\psi}\left(\frac{P}{E_a{}^2}\right) = \frac{\sqrt{\varepsilon_r}\,D^2}{480}\left(\frac{-2\ln\psi}{\psi^3} + \frac{1}{\psi^3}\right) = \frac{\sqrt{\varepsilon_r}\,D^2}{480}\cdot\frac{1-2\ln\psi}{\psi^3}$$

… so, equating the result to zero, $1 - 2\ln\psi = 0$ so $\ln\psi = 1/2$ and $\psi = \sqrt{e} = 1.649$. Thus it follows that, for some particular maximum field strength, the maximum power can be carried by a co-axial line when its $\psi = 1.649$ or $Z_0 = 30/\sqrt{\varepsilon_r}$ ohms.

Then $P = E_a{}^2\,0.00208D^2\sqrt{\varepsilon_r}$ so the absolute value of the maximum power increases with the square of the diameter of the co-axial cable as well as the square of the tolerable maximum field strength. Exceeding this maximum power would cause the dielectric strength of the medium between the conductors to be exceeded, leading to some kind of breakdown and a transient current path shunting the transmission line. This would probably give rise to a transient high VSWR which should 'trip' the transmitter feeding power into the line but, of course, this should never happen during normal operation. It is sometimes written that air-spaced lines can recover from 'flashover' but a high-current DC arc can cause migration of metal atoms between conductors (a kind of 'electro-plating' action) creating 'high spots' where the field strength could be even greater thereafter – I wonder if this happens with a UHF AC arc.

Dielectric breakdown is one of the two failure mechanisms that limit the maximum power that can be carried by a transmission line; the other is overheating on account of loss in the line. Obviously, failure on account of flash over will take a lot less time to act than overheating so, depending on the relative magnitudes of the two power limits, there is scope for the former and latter to apply to the short-duration-peak and mean powers, respectively, of a signal that changes in amplitude with time, like an RF television signal.

The next step is to calculate the loss in the transmission line but I will limit the treatment to the electrical characteristics; for consideration of the thermal characteristics I will make reference to a practical example later on.

6. Dealing with square roots by binomial expansion

However, before we go there I need to discuss briefly an arithmetic technique I will rely upon. As is becoming apparent, the topic of transmission lines is riddled with functions involving square roots and some conversion or approximation is needed if these are to be manipulated analytically. A common technique is to expand the square root into a series and then to truncate the series to the first two or three terms. Using a binomial series, in particular the Taylor series [9]:

$$(1+y)^n = \sum_{k=0}^{\infty} \frac{(-n)_k}{k!}(-y)^k = 1 + ny + \frac{1}{2}n(n-1)y^2 + \frac{1}{6}n(n-1)(n-2)y^3 + \ldots$$

… where $(-n)_k$ represents the 'falling factorial', the effect of which can be seen from the expansion.

Putting $n = ½$ for a square root and substituting $u = 1 + y$, so $y = u - 1$:

$\sqrt{u} \approx 1 + \frac{1}{2}(u-1) - \frac{1}{8}(u-1)^2$ taking the first three terms, or $\sqrt{u} \approx 1 + \frac{1}{2}(u-1)$ taking only the first two.

These are very convenient for turning a square root into something more manageable but the two term expansion is equivalent to representing a curve by a straight line from the origin and relies on the function inside the root having a small value, often given as <1. Where I will need to apply these, unfortunately, the function will always have a value ≥1 so the degree of approximation will be significant for some values of the function.

In practice, u is a polynomial with a constant term a as well as terms in bx and cx^2 so taking the first three terms of the expansion yields:

$$\begin{aligned}
\sqrt{u} &= 1 + \tfrac{1}{2}(a - 1 + bx + cx^2) - \tfrac{1}{8}(a - 1 + bx + cx^2)^2 \\
&= \tfrac{1}{2} + \tfrac{1}{2}a + \tfrac{1}{2}bx + \tfrac{1}{2}cx^2 \\
&\quad - \tfrac{1}{8}(a^2 - a + abx + acx^2 - a + 1 - bx - cx^2 + abx - bx + b^2x^2 + bcx^3 + acx^2 - cx^2 + bcx^3 + c^2x^4) \\
&= \tfrac{1}{2} - \tfrac{1}{8} + \tfrac{1}{2}a + \tfrac{1}{8}a + \tfrac{1}{8}a - \tfrac{1}{8}a^2 + \tfrac{1}{2}bx + \tfrac{1}{8}bx - \tfrac{1}{8}abx - \tfrac{1}{8}abx + \tfrac{1}{8}bx \\
&\quad + \tfrac{1}{2}cx^2 + \tfrac{1}{8}cx^2 - \tfrac{1}{8}acx^2 - \tfrac{1}{8}acx^2 - \tfrac{1}{8}b^2x^2 + \tfrac{1}{8}cx^2 - \tfrac{1}{8}bcx^3 - \tfrac{1}{8}bcx^3 - \tfrac{1}{8}c^2x^4 \\
&= \tfrac{3}{8} + \tfrac{3}{4}a - \tfrac{1}{8}a^2 + \tfrac{5}{8}bx - \tfrac{1}{4}abx + \tfrac{1}{8}bx + \tfrac{5}{8}cx^2 - \tfrac{1}{4}acx^2 - \tfrac{1}{8}b^2x^2 + \tfrac{1}{8}cx^2 - \tfrac{1}{4}bcx^3 - \tfrac{1}{8}c^2x^4
\end{aligned}$$

… so, neglecting terms in x^2 and above:

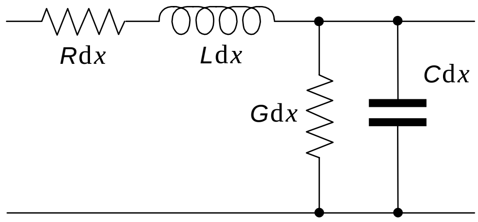
$$\sqrt{u} = \tfrac{3}{8} + \tfrac{3}{4}a - \tfrac{1}{8}a^2 + \tfrac{3}{4}bx - \tfrac{1}{4}abx$$

On the other hand, if only the first two terms of the expansion are taken and terms in x^2 and above are discarded:

$$\begin{aligned}
\sqrt{u} &= 1 + \tfrac{1}{2}(a - 1 + bx + cx^2) \\
&= \tfrac{1}{2} + \tfrac{1}{2}a + \tfrac{1}{2}bx
\end{aligned}$$

… which is rather different. Evidently, taking only the first two terms of the expansion can be a significant approximation. Indeed, taking any number of terms without comparing the outcome with a longer series introduces an approximation of unknown proportions!

7. Loss



The loss per unit length of a co-axial transmission line has two components: resistive loss in the surfaces of the conductors (increased by the 'skin depth' phenomenon); and dielectric loss in the medium between the conductors. These types of loss have the effect of introducing real, dissipative terms, R in series with the series inductance L, and G in parallel with the shunt capacitance C, respectively, which would be otherwise be purely reactive in a lossless line.

The full expression for the characteristic impedance of a transmission line in the presence of loss becomes:

$$Z_0 = \sqrt{\frac{R + \mathrm{j}\omega L}{G + \mathrm{j}\omega C}} \quad \text{ohms}$$

… where R is the series loss resistance per unit length ($\Omega \mathrm{m}^{-1}$), G is the shunt loss conductance per unit length (siemens m^{-1}), and ω is the angular frequency (radian s^{-1}) so, generally, Z_0 is complex.

When R and G are small (but still significant), Kraus tells us [10] this can be approximated as follows. I found a way to do this by re-arranging the expression, applying the binomial expansion and taking the first two terms:

$$Z_0 = \sqrt{\frac{L}{C}}\sqrt{\frac{R/\omega L + \mathrm{j}}{G/\omega C + \mathrm{j}}} \approx \sqrt{\frac{L}{C}}\left[1 + \frac{(R/2\omega L) - (G/2\omega C)}{\mathrm{j} + G/\omega C}\right]$$

… and then neglecting the $G/\omega C$ term in the denominator (… what a liberty, but little more so than the expansion, and it would most likely be small):

$$Z_0 \approx \sqrt{\frac{L}{C}}\left[1 + \mathrm{j}\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right]$$

Of course, Kraus might have used some other expansion that obscures the blatant omission.

If $G/C = R/L$, which is known as *Heaviside's condition* for a distortionless line, then $Z_0 = \sqrt{L/C}$ once again, which is purely real. One might imagine that trying to use a transmission line which imposes its own complex characteristic impedance on whatever it is connecting could make matching something of a hit-and-miss affair if, indeed, possible!

In this context, 'distortionless' appears to refer to an absence of dispersion in the transmission line, that is, variation of the velocity of propagation with frequency, which causes linear distortion. Different parts of a signal sent down the line simultaneously can arrive at different times and this effect tends to limit the maximum usable bandwidth. Oliver Heaviside's pioneering work on transmission lines in the 1880s [11] was directed initially towards improving telegraphy, where the shapes of on/off pulses must have become distorted when subjected to the significant dispersion of long transmission lines (i.e. pairs of telegraph wires). Later, he turned his attention to telephony and, particularly, to the new undersea cables. Heaviside is noted for first suggesting the addition of lumped inductors to a long transmission line, connected in series at intervals along its length to increase the distributed inductance (L) in order to counteract excessive distributed series resistance (R). His name is also associated with the step function that occurs at the beginning and end of each (dot or dash) symbol in simple telegraphy, and a layer in the ionosphere.

In a lossless line $G = R = 0$ and $Z_0 = \sqrt{L/C}$ again, but this also applies if the losses are not zero but $\omega L \gg R$ and $\omega C \gg G$, which is most likely in practice, and which certainly suits the above approximation.

The propagation constant from Section 3 becomes:

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

… where α and β are defined as before. Clearly from this expression, the significance of some absolute value of R or G depends on its size relative to L or C.

The forward wave defined in Section 3 for the lossless case becomes $V_2 e^{-\alpha x} e^{j(\omega t-\beta x)}$ which is attenuated exponentially with increasing distance x from the generator. The simplest way to express this attenuation is to use the unit neper m^{-1}. 1 neper corresponds to a voltage ratio of $1/e$ so a voltage ratio of α neper m^{-1} corresponds to the factor $e^{-\alpha}$.

Generally a voltage ratio v_1/v_2 can be expressed as $n = \ln(v_1/v_2)$ neper, whereas in decibels it is $20\log_{10}(v_1/v_2)$ so α can be converted to decibels per metre by taking $20\log_{10}e^{\alpha}$.

Rearranging the expression for γ and taking the first two terms of the binomial expansion again:

$$\gamma = j\omega\sqrt{LC}\sqrt{\left(1+\frac{R}{j\omega L}\right)\left(1+\frac{G}{j\omega C}\right)} = j\omega\sqrt{LC}\sqrt{1-\frac{j}{\omega}\left(\frac{R}{L}+\frac{G}{C}\right)-\frac{RG}{\omega^2 LC}} \approx j\omega\sqrt{LC}\left[1-\frac{j}{2\omega}\left(\frac{R}{L}+\frac{G}{C}\right)-\frac{1}{2\omega^2}\cdot\frac{RG}{LC}\right]$$

If $\omega L \gg R$ and $\omega C \gg G$, again, a useful substitution is:

$$\left(\frac{R}{\omega L}-\frac{G}{\omega C}\right)^2 = \left(\frac{R}{\omega L}\right)^2+\left(\frac{G}{\omega C}\right)^2-2\frac{RG}{\omega^2 LC} \quad \text{so} \quad \frac{RG}{LC} \approx -\frac{1}{2}\left(\frac{R}{L}-\frac{G}{C}\right)^2$$

Then:

$$\gamma \approx j\omega\sqrt{LC}\left[1-\frac{j}{2\omega}\left(\frac{R}{L}+\frac{G}{C}\right)+\frac{1}{4\omega^2}\left(\frac{R}{L}-\frac{G}{C}\right)^2\right]$$

… so $\alpha = \mathrm{Re}\,\gamma \approx \dfrac{\sqrt{LC}}{2}\left(\dfrac{R}{L}+\dfrac{G}{C}\right)$ and if Heaviside's condition is met, $G/C = R/L$ so $\alpha = \sqrt{LC}$.

In any case, multiplying through:

$$\alpha = \frac{1}{2}\left(\frac{R}{\sqrt{L/C}}+G\sqrt{L/C}\right) \text{ and substituting } Z_0 = \sqrt{L/C} \text{ gives } \alpha \approx \frac{1}{2}\left(\frac{R}{Z_0}+GZ_0\right)$$

… so $\alpha = 0$ for a lossless line ($R = G = 0$). Hereafter, I'll assume it's understood that this last expression is an *approximation* for α and I won't keep using the 'almost equals' sign, but it's probably close enough for the present context.

Again, if Heaviside's condition is met, $G = RC/L = R/Z_0^2$ and $\alpha = \dfrac{1}{2}\left(\dfrac{R}{Z_0}+\dfrac{R}{Z_0}\right)$

… so there would be equal contributions of loss from the series conductor resistance and the shunt dielectric conductance. This can be used as a test of whether or not Heaviside's condition is met.

I should note that some distinguished authors (e.g. [12]) say Heaviside's condition relates not to dispersion but to the attenuation constant α acquiring a frequency response on account of the condition not being met. This doesn't follow from the expressions above where ω cancels and later I will demonstrate the relation to dispersion. Also, reading Heaviside's own documentation [11] didn't help me to clarify this so, I regret, I will have to leave this unresolved.

7.1 Series conductor loss

For the moment, let's consider only series loss resistance (i.e. $G = 0$), so $\alpha = R/2Z_0$. In a co-axial cable there will be components of resistance from both conductors, R_a and R_b, (Ωm^{-1}) which will appear in series. The resistance R of a piece of material with a resistivity S (Ωm) and a simple shape, such as a cube, is given by $R = S\ell/A$ where ℓ (m) is the length of the current path through the piece of material and A (m^2) is the uniform cross-sectional area of the shape transverse to the direction of current flow. In the present context, the shape is the conducting surface of each conductor, so a pair of a cylinders, in each case having a thickness dependent on the 'skin effect'.

The skin effect [13] is the manifestation of the depth of penetration of an electromagnetic wave into the surface of a conducting material. In a transmission line, power is carried by an electromagnetic wave that propagates in the medium between the conductors. To some extent, the wave penetrates the surfaces of the conductors giving rise to alternating currents in them, but the magnetic fields created by these currents force the currents to flow predominantly in the surfaces of the conductors, not throughout their volume.

The magnitude of the current decays exponentially from the surface into the volume of each conductor and this can be represented by an equivalent current of uniform magnitude throughout the skin depth, d where:

$$d = \sqrt{\frac{S}{\pi\mu f}}$$

… where f is the frequency (Hz) and μ is the permeability ($\mathrm{Hm^{-1}}$) of the conducting material; $\mu = \mu_0\mu_r$ as before, where $\mu_0 = 400\pi$ $\mathrm{nHm^{-1}}$ and $\mu_r \approx 1$ for most non-ferromagnetic metals like copper. If it is assumed that both conductors of the co-axial line are made of the same material, which is frequently the case, the skin depth will be the same in them. $S = 17.5$ $\mathrm{n\Omega m^{-1}}$ for copper so:

$d =$ 2.11 μm at 1 GHz
6.66 μm at 100 MHz
21.1 μm at 10 MHz
66.6 μm at 1 MHz
0.21 mm at 100 kHz
0.66 mm at 10 kHz

Evidently, the actual thickness of either conductor in a practical co-axial cable will only affect the current distribution at low frequencies.

The cross-sectional area in which the current will flow in the inner conductor will be $\pi a^2 - \pi(a - d)^2 = 2\pi ad$, and in the outer conductor it will be $\pi(b + d)^2 - \pi b^2 = 2\pi bd$. Then, per unit length (i.e. $\ell = 1$), $R_a = S/2\pi ad$ and $R_b = S/2\pi bd$. The total resistance is given by:

$$R = R_a + R_b = \frac{S}{2\pi d}\left(\frac{1}{a}+\frac{1}{b}\right) = \frac{S}{\pi d}\cdot\frac{1+\psi}{D}$$

Substituting the expression for d, assuming $\mu_r = 1$ so $\mu = \mu_0 = 400\pi \times 10^{-9}$:

$$R = \sqrt{\frac{\mu_0 f S}{\pi}}\cdot\frac{1+\psi}{D} = \sqrt{400\times10^{-9} f S}\,\frac{1+\psi}{D}$$

The ratio of the resistances is equal to the reciprocal of the ratio of the radii, $R_a/R_b = b/a = \psi$, and the power dissipated in each conductor will be proportional to its resistance so the fraction of the total power dissipated in the inner conductor $P_a = b/(a + b) = 1/(1 + 1/\psi)$.

Considering the attenuation constant, α:

$$\alpha = \frac{R}{2Z_0} = \frac{\sqrt{400\times10^{-9} f S}}{2Z_0}\cdot\frac{1+\psi}{D} = \sqrt{10^{-7} f S}\,\frac{1+\psi}{DZ_0} \quad (\text{neper m}^{-1})$$

… then substituting the expression that relates Z_0 to the cable geometry for a lossless line, which amounts to a further approximation but no greater than the liberties already taken!

$$\alpha = \frac{\sqrt{10^{-7} f S\varepsilon_r}}{60}\cdot\frac{1+\psi}{D\ln\psi} = 5.27\times10^{-6}\,\frac{\sqrt{f S\varepsilon_r}}{D}\cdot\frac{1+\psi}{\ln\psi}$$

If D is held constant and ψ is varied over $1.05 \le \psi \le 10$, the value of the variable part of this expression will vary from 42.01 to 4.78 via a range of smaller numbers such as 3.73 at $\psi = 5$ so it must pass through a minimum which, once again, can be found by equating the differential to zero.

Omitting the constant multiplier $5.27\times10^{-6}\sqrt{f S\varepsilon_r}\big/D$ and using $\dfrac{\mathrm{d}}{\mathrm{d}x}\dfrac{u}{v} = \dfrac{1}{v^2}\left(v\dfrac{\mathrm{d}u}{\mathrm{d}x} - u\dfrac{\mathrm{d}v}{\mathrm{d}x}\right)$:

$$\frac{\mathrm{d}\alpha'}{\mathrm{d}\psi} = \frac{1}{(\ln\psi)^2}\left[(\ln\psi)\frac{\mathrm{d}}{\mathrm{d}\psi}(1+\psi) - (1+\psi)\frac{\mathrm{d}}{\mathrm{d}\psi}\ln\psi\right] = \frac{(\ln\psi) - (1+\psi)/\psi}{(\ln\psi)^2} = 0$$

… so $\ln\psi = 1 + 1/\psi$.

The solution to this is found in the 'Lambert W-function' [14, also known by the names 'ProductLog' and 'Omega function'], $\mathrm{W}(x)$, which satisfies the relation $\ln \mathrm{W}(x) = \ln x - \mathrm{W}(x)$. Making the substitution $\psi = 1/\mathrm{W}(x)$ so $\ln\psi = -\ln \mathrm{W}(x)$:

$$-\ln \mathrm{W}(x) = 1 + \mathrm{W}(x)$$

… but according to the relation: $-\ln \mathrm{W}(x) = -\ln(x) + \mathrm{W}(x)$

… so: $\ln(x) = -1$ and $x = 1/e$

From the series expansion:

$$W(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1} n^{n-2}}{(n-1)!} x^n = x - x^2 + \frac{3}{2}x^3 - \frac{8}{3}x^4 + \frac{125}{24}x^5 - \frac{54}{5}x^6 + \frac{16807}{720}x^7 - \ldots$$

… it appears that $\mathrm{W}(1/e) = 0.2784$ although, when I calculated this with MS Excel, the fourth significant figure was still converging after 144 terms which reached the limit of large- and small-number handling in Excel for my simple implementation of the series.

So $\psi = 3.591$, and with this value $(1 + \psi)/\ln\psi = 3.591$ as well!

It follows that for lowest conductor loss $Z_0 = 76.7/\sqrt{\varepsilon_r}$ ohms.

Then $\alpha = 18.93\times10^{-6}\sqrt{f S \varepsilon_r}\big/D$ (neper m^{-1}) so this component of loss is proportional to the reciprocal of the diameter of the transmission line and to the square-root of the frequency.

In view of the proportionality of α to R it also follows that the proportion of the total conductor loss that can be attributed to the surface of the inner conductor is $1/(1 + 1/\psi)$ which, obviously, is always greater than ½ for realistic values of ψ.

The working in Section 4 that led to an expression for the distributed inductance, and then to the characteristic impedance, was based on the tacit assumption that the skin depth was zero and the current was flowing in the facing surfaces of the co-axial conductors: the integral was performed over the range $a \le r \le b$. At low frequencies, current flowing within the finite skin depth will increase the range of integration, potentially increasing the distributed inductance, so the characteristic impedance would be expected to rise as the frequency falls, to some small extent. This effect doesn't appear to be well documented but I did notice reference to it in an article [15] published by a manufacturer of co-axial cables. This may be significant in high-power short-wave transmitting stations.

7.2 Shunt dielectric loss

Now, let's consider only dielectric loss by putting $R = 0$ in:

$$\alpha \approx \frac{1}{2}\left(\frac{R}{Z_0} + GZ_0\right)$$

… so $\alpha = GZ_0/2$ and it is more or less intuitive that any value of loss conductance would have greater effect when shunting a higher impedance line.

Loss in a dielectric material implies that its relative permittivity is complex rather than being a purely real dimensionless number, and this is usually represented by the 'loss tangent' $\tan\delta$; the ratio of the imaginary to real parts of ε_r which is also known as the 'power factor':

$$\tan\delta = -\frac{\mathrm{Im}\,\varepsilon}{\mathrm{Re}\,\varepsilon} = \frac{\sigma}{\omega\varepsilon}$$

… where $\varepsilon = \varepsilon_0\varepsilon_r$ as usual, and σ is the effective conductivity of the material. For polythene at 100 MHz, $\varepsilon_r = 2.26$ so $\varepsilon = 20.0$ pFm^{-1} and $\tan\delta = 0.0002$ so $\sigma = 2.51$ μsiemens m^{-1}. The dielectric properties of polythene appear to be more-or-less constant throughout the frequency range up to about 1 GHz and if $\tan\delta$ remains constant, σ must be proportional to the frequency: $\sigma = 2\pi f\varepsilon_0\varepsilon_r\tan\delta = 55.61\times10^{-12} f\varepsilon_r\tan\delta$.

Incidentally, I avoided using the symbol δ to represent the skin depth although this appears customary in the well-known text books. In this context, using δ for two different parameters both related to attenuation could easily provoke a muddle.

I should mention here that although it is convenient for calculation to represent dielectric loss by an effective conductivity, in most materials with good insulating properties that would be considered for use in a transmission line the actual mechanism for the loss is what is known as 'dielectric hysteresis' [16]. This appears to be poorly explained in the well-known text books on electromagnetics, even though the process has become almost as common as electric light (since the development of the microwave oven). What is clear is that the applied electric field causes polarisation of atoms or molecules, distorting the configuration of their nuclei and the electron clouds around them, and this distortion changes in magnitude and direction in sympathy with the alternating field. The relationship between the field strength and the degree of polarisation can be depicted by an ellipse – the latter always lags the former because there are moving parts (albeit very tiny ones, with lots of tiny masses that need to be accelerated and decelerated).

Then somehow, a fraction of the energy in the resulting periodic atomic-scale vibration gets coupled into random, aperiodic agitation of adjacent atoms or molecules and this energy is irreversibly lost as heat. It may not be obvious that some of the population of atoms or molecules vibrating in phase with each other should 'bump' into one another (things probably never touch on an atomic scale), although people on a dance floor moving to the same beat sometimes bump into one another – there may be some degree of randomness in their individual interpretations of the ideal dance sequence. At the atomic scale we're led to believe all the electrons are spinning around their nuclei in probabilistic orbitals, so perhaps it follows that a proportion would collide.

So dielectric loss converts electromagnetic energy to heat, which is lost from the transmission line by the usual mechanisms: conduction, convection and infra-red radiation. In addition, there can be some degree of non-linearity in this process, rather like that encountered in magnetisation, hence the use of the name 'hysteresis'. We're told that some materials exhibit appreciable retention of electrical polarisation from one part of the cycle to another; a short-term manifestation of the 'electret' principle. Then, the relationship between the degree of polarisation and the field strength can be depicted by the well-known hysteresis curve (like a curved parallelogram), demonstrating substantial non-linearity and probably additional loss. We're also told of resonances, where the angular frequency of the applied field equals the angular frequency of the electrons in one of the orbitals, or a harmonic of it.

Evidently, this is a substantial area of physics in its own right and perhaps the electromagnetics generalists can be excused for their light touch.

The conductance G of a piece of material with a conductivity σ and a simple shape, such as a cube, is given by $G = \sigma A / \ell$, where A is the cross-sectional area of the shape, transverse to the direction of current flow, and ℓ is the length of the current path through the piece of material. In the present context, the path of the effective current would follow the lines of electric force, radially out from the inner conductor to the outer one. Then the cross-sectional area increases with radial distance r towards the outer conductor, like a wedge. It seems reasonable to think of successive cylindrical shells surrounding the inner conductor, and one another, in order to find the result by integration, but the parameter that would add is the effective 'resistance' of the dielectric medium; the reciprocal of its effective conductance, $1/G = \ell/\sigma A$.

Per unit length of line $A_{(r)} = 2\pi r$, $a \le r \le b$ and $\ell = \mathrm{d}r$, then:

$$\frac{1}{G} = \int_a^b \frac{\mathrm{d}r}{2\pi\sigma r} = \frac{1}{2\pi\sigma}\int_a^b \frac{1}{r}\mathrm{d}r = \frac{1}{2\pi\sigma}\ln\psi \quad \text{so} \quad G = 2\pi\sigma\frac{1}{\ln\psi}$$

… and then:

$$\alpha = \frac{GZ_0}{2} = \frac{60\pi\sigma}{\sqrt{\varepsilon_r}}\cdot\frac{\ln\psi}{\ln\psi} = \frac{60\pi\sigma}{\sqrt{\varepsilon_r}} \quad (\text{neper m}^{-1})$$

… so α is independent of the geometry of the transmission line and depends only on the characteristics of the dielectric material. In keeping with the 'conduction' metaphor, it would follow that increasing the impedance of the line would create a longer current path through the lossy dielectric medium, reducing the conductance, so the product of G and Z_0 would remain constant.

Substituting for $\sigma = 55.61\times10^{-12} f\varepsilon_r\tan\delta$:

$$\alpha = 10.48\times10^{-9} f\tan\delta\sqrt{\varepsilon_r} \ (\text{neper m}^{-1})$$

… so the dielectric loss is proportional to the frequency.

7.3 Series and shunt losses together

To complete the picture, we can now put together both types of loss in:

$$\alpha = \frac{R}{2Z_0} + \frac{GZ_0}{2}$$

Inserting the expressions from the previous two sections:

$$\alpha = 5.27\times10^{-6}\sqrt{S\varepsilon_r}\,\frac{\sqrt{f}}{D}\cdot\frac{1+\psi}{\ln\psi} + 10.48\times10^{-9} f \tan\delta\sqrt{\varepsilon_r} \quad \text{or} \quad \alpha = k_1\frac{\sqrt{f}}{D}\cdot\frac{1+\psi}{\ln\psi} + k_2 f$$

… where $k_1 = 5.27\times10^{-6}\sqrt{S\varepsilon_r}$ and $k_2 = 10.48\times10^{-9}\tan\delta\sqrt{\varepsilon_r}$.

Taking typical values for a co-axial cable that has both losses, for example one with copper conductors and solid polythene dielectric: $S = 17.5\times10^{-9}\ \Omega\text{m}^{-1}$, $\varepsilon_r = 2.26$ and $\tan\delta = 0.0002$ so $k_1 = 1.048\times10^{-9}$ and $k_2 = 3.151\times10^{-12}$ in appropriate units (whatever they might be!).

Polythene dielectric is generally used in order to make the cable flexible and the largest diameter I have seen listed [17] for a 50 Ω co-axial cable with solid polythene dielectric is D = 23.1 mm for RG-20 cable which, apparently, has the modern designation M17/81-00002. Apart from the obvious loss of flexibility, I will explain the electrical upper limit on D in the next section and how it is related to the maximum frequency at which a cable can operate.

Incidentally, because of the presence of the logarithm in the expression relating the characteristic impedance to the geometry, it is impractical to obtain values of Z_0 much greater than a few hundred ohms. Even for an air-filled conducting tube the size of a London Underground tunnel (3.7 m diameter), a 1 mm diameter wire at the centre of would yield only Z_0 = 493 Ω (my thanks to John Sykes of BBC World Service for this observation).

Returning to RG-20 cable, taking D = 0.023 m and ψ = 3.50 that yields 50 Ω characteristic impedance for solid polythene dielectric:

$$\alpha = 3.592 k_1 \sqrt{f}/D + k_2 f = 163.6\times10^{-9}\sqrt{f} + 3.151\times10^{-12} f \quad (\text{neper m}^{-1})$$

Inserting values over a range of frequencies:

at f = 1 GHz,	$\alpha = 5.2\times10^{-3} + 3.15\times10^{-3}$	or 7.3 dB / 100 m
at f = 100 MHz,	$\alpha = 1.64\times10^{-3} + 0.32\times10^{-3}$	or 1.7 dB / 100 m
at f = 10 MHz,	$\alpha = 0.52\times10^{-3} + 0.032\times10^{-3}$	or 0.48 dB / 100 m
at f = 1 MHz,	$\alpha = 0.16\times10^{-3} + 0.0032\times10^{-3}$	or 0.14 dB / 100 m
at f = 100 kHz,	$\alpha = 0.052\times10^{-3} + 0.00032\times10^{-3}$	or 0.045 dB / 100 m
at f = 10 kHz,	$\alpha = 0.016\times10^{-3} + 0.000032\times10^{-3}$	or 0.014 dB / 100 m

…. where the first part of α represents the conductor loss and the second part the dielectric loss. The attenuation of RG-20 cable is given [17, Page 40] as 0.6 dB / 100 feet at 100 MHz which corresponds to 1.97 dB / 100 m; the value above (1.7 dB) is fairly close.

From the expression for C given in Section 4, $C = 2\pi\varepsilon/\ln\psi$, a value of 100.3 pF m^{-1} would be expected for polythene dielectric and ψ = 3.50. However, for RG-20 and several other 50 Ω flexible co-axial cables the catalogue value is 101.1 pF m^{-1} implying a slightly different combination of ε_r and ψ. Perhaps this has something to do with the use of stranded conductors. From the expression for Z_0 given in Section 4 (for a lossless line), L = 250.8 nHm^{-1}.

Also, putting the above (calculated) attenuation values for RG-20 at 100 MHz ($\omega = 628.3\times10^{6}$ rad s^{-1}) into the expression for α at the beginning of this section, R = 0.164 Ωm^{-1} and $G = 12.8\times10^{-6}$ Sm^{-1}.

Therefore ωL = 157.6 which is much larger than R, and ωC = 0.063 which is much larger than G, so the approximations in Section 7 are justified, as is the use of the lossless-line expression for Z_0 a moment ago. These inequalities still apply at lower frequencies, although at 10 kHz ωL is only about ten times the magnitude of R (because of the $\sqrt{f}$ relationship on account of skin depth).

This is all for the largest diameter flexible co-axial cable I could find but what about a more-common polythene-dielectric cable like the 5 mm (outside) diameter RG-58? In this case $D = 2.95\times10^{-3}$ m and taking $\psi = 3.50$ again, $\alpha = 1.276\times10^{-6}\sqrt{f} + 3.151\times10^{-12}f$ (neper m^{-1}) so at 100 MHz, $\alpha = 12.8\times10^{-3} + 0.32\times10^{-3}$ or 11.4 dB / 100 m. In this case comparison with a catalogue value [17] is not so close; the typical value is given as 4.6 dB / 100 feet which corresponds to 15.1 dB / 100 m. Nevertheless, the derived values are $R = 1.28\ \Omega\text{m}^{-1}$ and $G = 12.8\times10^{-6}\ \text{Sm}^{-1}$, while the values of C and L are the same as before, so the same conclusion would be reached.

Below about 1 GHz, then, it can be concluded that conductor loss is consistently greater than dielectric loss – more so for smaller cable diameters. This may be borne out in [17] where 'loss constants' are listed, unfortunately without explanation or a unit (and I haven't been able to find these in the, extensive, US military MIL-C-17 specification [18]). The values given for RG-20 cable are 0.052 and 0.00126 for resistive and dielectric loss, respectively, and if these are in consistent units they probably support this conclusion. All the values given for other cables follow suit.

At around 1 GHz, and above, the dielectric loss of RG-20 cable exceeds the conductor loss but then solid polythene dielectric would not be appropriate. Solid PTFE is popular for microwave frequencies; its different $\varepsilon_r = 2.1$ will affect both components of loss equally and at 3 GHz its $\tan\delta = 0.00015$ is better than that of polythene (0.00031). Inserting these values for a 50 Ω copper / PTFE cable ($\psi = 3.353$ in this case) with $D = 5$ mm, working at $f = 3$ GHz:

$$\alpha = 0.04 + 0.007 \text{ (neper m}^{-1}\text{)}$$

… which corresponds to about 0.4 dBm^{-1} (or 40 dB / 100 m!).

Thus it appears generally true that conductor loss is greater than dielectric loss in co-axial cables at frequencies below SHF and, because the dielectric loss is not affected by the geometry, the minimum combined loss occurs at the same value of ψ as the minimum conductor loss. At SHF and above it appears that dielectric loss could become dominant but the relatively large absolute magnitude of the combined loss makes other, more-complicated, but lower-loss forms of transmission line preferable for long runs, such as waveguides.

8. Phase constant, phase and group velocities, dispersion, velocity factor, waveguide modes

Now returning to the imaginary part of the propagation constant expressed in Section 7; for a lossy transmission line, the propagation phase constant becomes:

$$\beta = \operatorname{Im}\gamma \approx \omega\sqrt{LC}\left[1 + \frac{1}{4\omega^2}\left(\frac{R}{L} - \frac{G}{C}\right)^2\right]$$

… and if Heaviside's condition is met, $G/C = R/L$ so $\beta = \omega\sqrt{LC}$, which would also apply if the line were lossless.

The velocity of propagation v through the transmission line is given by:

$$v = \frac{\omega}{\beta} = \frac{1}{\beta/\omega} = \frac{1}{\sqrt{LC}\left[1 + \frac{1}{4\omega^2}\left(\frac{R}{L} - \frac{G}{C}\right)^2\right]}$$

… which has a constant value of $1/\sqrt{LC}$ if Heaviside's condition is met or the line is lossless. Otherwise the velocity depends on the frequency so the line exhibits dispersion. From this expression, the velocity will be smaller than the non-dispersive case at low frequencies, tending towards $1/\sqrt{LC}$ as the frequency increases. However, the presence of the squared frequency in the denominator suggests this mechanism for dispersion is significant only at low frequencies.

Sometimes it is useful to distinguish [19] between the:

- *phase velocity* v_p which indicates how rapidly some particular point in an unchanging travelling wave travels (e.g. the positive-going zero crossing), defined as $v_p = \omega/\beta$ as above; and the …
- *group velocity* v_g at which energy propagates; defined as $v_g = d\omega/d\beta$.

The latter is always smaller than the ‘speed of light’ $c \approx 300\times10^{6}\ \mathrm{ms}^{-1}$ but in some circumstances, like waveguide propagation, the phase velocity can exceed c. Unfortunately modulation travels at the same rate as energy, or changes in the presence, amplitude or phase of some wave that conveys energy, so communication remains resolutely sub-luminal! In a dispersive medium different group velocities are found at different frequencies and the phase and group velocities may differ.

For our transmission line, v_p is as given before, whilst $v_g = \mathrm{d}\omega / \mathrm{d}\beta$. For a uniform line, the *group delay* is the reciprocal $\mathrm{d}\beta / \mathrm{d}\omega$, given by:

$$\frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \sqrt{LC}\,\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega + \frac{1}{4\omega}\left(\frac{R}{L}-\frac{G}{C}\right)^2\right] = \sqrt{LC}\left[1 - \frac{1}{4\omega^2}\left(\frac{R}{L}-\frac{G}{C}\right)^2\right]$$

The term containing $1/\omega^2$ represents the dispersion but I can’t properly account for why this appears to *reduce* the group delay at low frequencies, potentially making it negative, when the reduced velocity would suggest an increasing delay … time machine anyone?

The only excuse I can offer is that group delay variation by this mechanism could be termed a ‘second order’ effect, involving terms that are often omitted from text books because they are considered vanishingly small. In combination with the degree of approximation I set out in Section 6, perhaps this absurdity is hardly surprising.

Alternatively, avoiding the ‘useful substitution’ in Section 7:

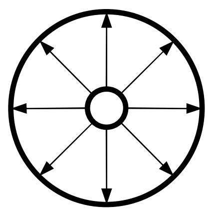
$$\beta \approx \omega\sqrt{LC}\left[1 - \frac{1}{2\omega^2}\cdot\frac{RG}{LC}\right] \quad \text{so} \quad \frac{\mathrm{d}\beta}{\mathrm{d}\omega} = \sqrt{LC}\,\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega - \frac{1}{2\omega}\cdot\frac{RG}{LC}\right] = \sqrt{LC}\left[1 + \frac{1}{2\omega^2}\cdot\frac{RG}{LC}\right]$$

… which is perhaps more believable in the context of delay, though not velocity.

Also (whilst confessing!), although I’ve given the expression for β here with a ‘4’ in the denominator, an old book by Schelkunoff [20] gives this with an ‘8’ instead, and I strongly suspect Schelkunoff is correct although I haven’t managed to achieve ‘8’ by my own efforts. Could he have done something clever with the third term of the expansion?

When an electromagnetic wave of any frequency propagates in free space, its group and phase velocities are equal to the ‘speed of light’, $c = 1/\sqrt{\mu_0\varepsilon_0}$. The mode of propagation distant from the source is a pure TEM (Transverse Electro-Magnetic) wave, also known as a plane wave, in which uniform electric and magnetic fields are found in the plane transverse to the direction of propagation. Because the velocity is constant with changing frequency, free space is considered non-dispersive.

The same TEM mode of propagation applies in any unbounded medium other than free space. It is an observed fact that the velocity of propagation v is lower than c and the relative velocity $v/c = 1/\sqrt{\mu_r\varepsilon_r}$ which is equal to $1/\sqrt{\varepsilon_r}$ in a dielectric medium not possessing significant magnetic properties. In the context of optics, $\sqrt{\varepsilon_r}$ is known as the ‘refractive index’ or ‘index of refraction’. The lower the velocity, the shorter the wavelength at a given frequency: $\lambda = v/f = c/f\sqrt{\varepsilon_r}$.



This applies equally to a wave propagating along a co-axial transmission line so long as the radial dimensions are much smaller than λ. The principal mode of propagation is a TEM wave with cylindrical geometry having a radial E-field between the inner and outer conductors, like the electrostatic geometry I described at the outset and illustrated to the right. With larger relative dimensions, some of the power input to the line could be conveyed by waveguide modes travelling in the hollow tube that is the outer co-axial conductor. Waveguide propagation is naturally highly dispersive – for a given ‘guide geometry, the phase and group velocities both depend on the wavelength in the medium that fills the ‘guide which depends on the frequency. On the other hand, the TEM mode in air, in most common dielectric materials and normally in co-axial cables (apart from the small dispersion just described), is non-dispersive.

When waveguides are used for communication purposes their dispersiveness is rarely an issue because the fractional bandwidth in use is usually small on account of the attendant high UHF or SHF centre frequency. Group delay variation can be an issue in radar applications though.

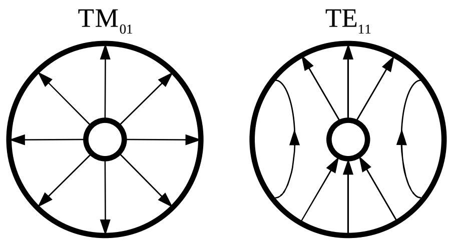
Because waveguide modes generally have different velocities of propagation from one another, and from a TEM wave, any mixture in a co-axial cable would correspond to multipath propagation which is generally to be avoided (*viz* delayed-image interference impairs analogue television and multipath-at-source can compromise the performance of DTT).

Waveguide-mode propagation between co-axial cylindrical conductors separated by a dielectric medium is subject to a cut-off wavelength λ_C that depends on the radii of the conductors, the characteristics of the dielectric and the particular waveguide mode. If the dielectric is air, when the air-wavelength $\lambda_0 = c/f$ is longer than λ_C the wave in the 'guide is attenuated rather than transmitted. The attenuation constant in this case is given by [21]:

$$\alpha = \frac{2\pi}{\lambda_0}\sqrt{\left(\frac{\lambda_0}{\lambda_c}\right)^2 - 1} \text{ (neper m}^{-1}\text{)}$$

With some other dielectric medium, as before, these wavelengths would be divided by $\sqrt{\varepsilon_r}$. Thus, to avoid significant power being carried in waveguide modes, co-axial cables are usually operated at frequencies for which λ_0 exceeds λ_C by a significant ratio or, to put it another way, the dimensions of the cable are chosen to fulfil this condition at the highest working frequency.

The TM_{01} co-axial waveguide mode would naturally couple efficiently from the co-axial TEM wave because it has a similar arrangement of (radial) electric and (perpendicular, axial) magnetic fields, and $\lambda_C = 2.029(b - a)/\sqrt{\varepsilon_r}$ for this mode [22]. However, the TE_{11} mode, in which the distribution of the radial electric field has a peak at some particular 'polarisation', has the longest $\lambda_C = 2.950(b + a)/\sqrt{\varepsilon_r}$ (note the sum, rather than difference in this case) so it is usually considered the 'dominant mode' for this type of waveguide – it can carry power at frequencies for which all other modes are cut off. I've sketched the arrangements of the E-field for these two modes to the right.



For example, to achieve an attenuation of >20 dBm^{-1} for the TE_{11} mode in an air-dielectric cable operating at a maximum frequency of 1 GHz (α = 2.3 nepers m^{-1} and λ = 0.3 m), $b + a < 100$ mm which, taking into account the $\sqrt{\varepsilon_r}$ of a suitable dielectric, would easily be achieved for a flexible cable. At a maximum frequency of 10 GHz, $b + a < 10$ mm, and so on (λ_C tends towards λ_0 as they become « 1), which explains why co-axial cables for high microwave frequencies are usually made with small diameters. However, this can impose a frequency limitation on high-power air-spaced cables which need large diameters in order to withstand high voltages and / or to constrain the conductor loss.

Coupling of power from the TEM wave into the TE_{11} mode would most likely be caused where there was an irregularity in the shape of the outer conductor (e.g. a small dent or if it were squashed), or an asymmetry in the co-axial structure (e.g. displacement of the inner conductor). Something has to establish the preferred orientation of the 'polarised' electric field.

9. Summary and inferences

The principal mode of propagation in a uniform co-axial cable, a radial TEM travelling wave, does not have an upper cut-off frequency.

The greatest electric-field strength between the co-axial conductors occurs immediately adjacent to the outer surface of the inner conductor and this is minimised when $Z_0 = 60$ ohms for an air-spaced line, generally when $Z_0 = 60/\sqrt{\varepsilon_r}$ ohms, and then $E_a = 2Ve/D$ (e = 2.718) so the field strength is reduced when the diameter of the transmission line is increased.

However, if the limiting factor is the electric-field strength, the greatest power can be carried by a co-axial cable when $Z_0 = 30$ ohms for an air-spaced line, generally $Z_0 = 30/\sqrt{\varepsilon_r}$ ohms, and then $P = E_a^2\, 0.00208 D^2 \sqrt{\varepsilon_r}$ so the maximum power increases with the square of the diameter of the co-axial cable as well as the square of the tolerable field strength.

The smallest conductor loss is achieved when $Z_0 = 76.7$ ohms for an air-spaced line, generally when $Z_0 = 76.7/\sqrt{\varepsilon_r}$ ohms, and then $\alpha = 18.93\times10^{-6}\sqrt{f S \varepsilon_r}\,/D$ (neper m^{-1}) so this loss is reduced when the diameter of the transmission line is increased. This loss, because of the skin depth, is proportional to the square-root of the frequency. The skin depth in copper is measured in microns for radio frequencies down to about 1 MHz.

The dielectric loss is dependent on the characteristics of the dielectric material between the conductors but not on the diameters of the conductors. If $\tan\delta$ for the dielectric material is constant with changing frequency, which is the case for polythene, this component of loss is proportional to the frequency: $\alpha = 10.48\times10^{-9} f \tan\delta\sqrt{\varepsilon_r}$ (neper m^{-1}).

For the example of a flexible co-axial cable with copper conductors and solid polythene dielectric (probably the most common combination for flexible cables) working at any frequency for which that construction would be appropriate, the conductor loss dominates the combined loss. Dielectric loss can become comparable to conductor loss at SHF, and other dielectric materials are then used to provide smaller combined loss.

Whilst in many cases Heaviside's condition may not be met, in practice $\omega L \gg R$ and $\omega C \gg G$ so the imaginary part of the complex characteristic impedance is relatively small and doesn't appear to present matching difficulties.

Throughout its normal operating frequency range, a co-axial cable is non-dispersive and the velocity of propagation, relative to the speed of light, $v/c = 1/\sqrt{\varepsilon_r}$. However, at particularly low frequencies it can exhibit dispersion although the magnitude of this effect is difficult to estimate by approximate theory ;o)

Also, at high frequencies, where the outer conductor begins to behave as a hollow-tube waveguide, power can be conveyed by a highly-dispersive waveguide mode. In order to suppress waveguide propagation, the diameter of the outer conducting surface must be smaller than the TE_{11} cut-off wavelength $101.7\times10^{6}/f$ (metre) and preferably a small fraction of it (less than half).

10. Practical examples

10.1 Feeder at a high-power television transmitting station

Presently (2005) the most powerful analogue television transmitting stations in the UK use air-spaced 50 Ω co-axial cables of 6⅛" nominal diameter and typical specifications for such cable can be found in on-line catalogues such as [23]. In this case, for their 'HCA618-50JT Heliflex' cable, $Z_0 = 50\ \Omega$ and $a = 33.5$ mm. The external diameter of the outer conductor is given as 162 mm whilst 6⅛" corresponds to 155.6 mm, so the 6⅛" must refer to the inside diameter of the outer conductor, D, and the difference must account for the thickness of this conductor.

Although the dielectric is principally air, a plastic spiral spacer is used to keep the inner conductor in the centre and the 'velocity' is given as 97%, so the effective $\varepsilon_r = 1.063$. From the formula for the characteristic impedance it then follows that $\psi = 2.36$ which would make $D = 158.2$ mm. The resolution of this apparent discrepancy probably lies in the fact that both inner and outer conductors have a helical indentation which eases bending of the cable and, whilst ψ and Z_0 are associated with average dimensions, some of the specifications are probably maximum or minimum dimensions (also, 6⅛" is probably only a *nominal* figure). I will assume *effective* values of $a = 33.0$ mm and $D = 155.8$ mm.

From the formula for the maximum power handling on account of dielectric breakdown, $P = 8.038\times10^{-6}\,E_a^{2}$. For clean dry air between large, perfectly flat conducing plates, $E_a \approx 3$ kV/mm. In the case of this cable, the conductors are curved on account of being cylindrical and they have the helical indentations, the cable may be bent in some places to the specified [23] minimum bending radius, and the presence of the plastic spacer with greater ε_r than air will cause a local increase in the electric field strength. All these factors contribute to a smaller safe value for E_a. For the example cable, the specified peak power rating is 2.89 MW implying a maximum $E_a \approx 600$ V/mm and a peak sinusoidal voltage of 17 kV, which is also the specified value. Evidently, the manufacturer applies a 5× safety factor to the maximum electric-field strength relative to flat plates.

There are 18 television transmitting stations in the UK that radiate analogue television with an ERP of 500 kW and above for each of the four national services. Most of these have two co-axial feeders each connecting one of the two final combiners (adding the two BBC services to ITV and Channel 4) to half of the antenna system. The power per-service in each feeder ranges up to about 25 kW so the total power in each feeder is up to about 100 kW. It's 40 kW per service at Crystal Palace, but four feeders are used there feeding separate BBC and commercial television antennas, and the final combiner is the air! Each of these powers is the so-called 'peak sync.' value; that is, the peak-envelope power during each sync. pulse, and the sync. pulses for the different services may coincide. The analogue and digital (NICAM) sound carriers also add a little.

For the example cable (which is undoubtedly typical, although this particular brand may not be used at these stations), at 100 kW there would be a substantial safety margin, about 29× in terms of maximum power on account of dielectric breakdown, which would need to accommodate any standing waves on the line, any imperfections in its geometry and any contamination of the air dielectric (it is filtered and dried before it is pumped in, by a device that resides in a room known affectionately as the 'beer engine').

In the UK, each numbered channel in Band IV / V extends from 470 + 8 (channel number - 21) MHz to a frequency 8 MHz higher. The analogue vision carrier is 1.25 MHz from the lower edge of the channel and a DTT signal is centred in the channel. Channels 21 to 38 are known as Band IV and channels 39 to 68 are known as Band V. Channel 36 has been used for airfield radar and for connecting VCRs to television receivers, and Channel 38 is reserved for radio astronomy – neither is used for television broadcasting at the moment.

The highest-frequency channel used at the Durris station in Scotland, for example, for analogue television is Channel 32 which puts the vision carrier at 559.25 MHz. The average-power specification for the example cable is 93.5 kW when operated at 600 MHz and this is most likely based on heating caused by conductor loss. The loss is specified as 0.546 dB / 100 m at 600 MHz and the analogue antennas at Durris are at an average height of 312 m up the mast so there would be some 1.7 dB loss if this particular cable were used. If the average power input to the cable were 49 kW (i.e. roughly, all four services at black level) then about 16 kW would be dissipated in each of the two cables. This might sound a lot, but at less than 50 W / metre and considering the wind speeds encountered at high altitudes, it may be barely enough to prevent ice forming on the feeders in the winter. Incidentally, the actual average power of 49 kW is just over half the specified maximum 93.5 kW, and the likelihood of all four services going to black-level simultaneously is probably small nowadays. Also, this indicates that for analogue television the power ratings of feeders are, quite rightly, based on the average transmitter power, not the short-duration peak power.

The fraction of this power dissipated in the inner conductor $P_a = 1 / (1 + 1 / \psi) = 70\%$ of the total, that is about 11 kW and, somehow, all this heat gets to the outside. The air inside the feeders is static at these high-power stations. For the example cable, the spiral spacer is made from PTFE and the cable is rated for operation with an impressive maximum inner conductor temperature of 150° C!

Since the instantaneous-peak power limitation for this cable is 2.89 MW while its average power handling is specified as 93.5 kW, it could be said that the cable has an inherent *peak-to-mean ratio* of about 14.9 dB (or 31×). That is, it could operate just within specification when presented with an input signal of 93.5 kW average power having occasional peaks of short duration 14.9 dB greater in power.

10.1.1 What this means in the context of television switch-over

The COFDM generator can be likened to a 'fruit machine'; a 'one-arm bandit'. The phases of the multiple carriers, from symbol-to-symbol, convey the bits that make up the data stream. At the beginning of each new symbol, the handle is pulled down and the reels come up ACBA, or whatever (e.g. cherry, bar, dollar, cherry). That is, the carriers are switched to phases represented by …ACBA…, all 1705 of them in the present '2k' mode. With the intended future 64QAM modulation scheme, the 64 possible constellation points correspond to 52 distinct phases ($n = 64$; $m = n - 2\sqrt{n} + 4$) and the carriers take on phases from this 'alphabet' more-or-less randomly from symbol to symbol. So, at the beginning of any new symbol, there's a high probability that, say, 33 of the carriers will take on phase 'A', 33 of them phase 'B', etc., and with about 33 carriers or about 1.9% of them in each of the 52 possible phase states this would correspond to an even distribution of carriers in phases. The vector sum of all the carrier voltages would tend to zero and the power of the whole DTT signal would be the sum of the powers of the individual carriers, that is, 32 dB greater than the power of an individual carrier. Of course, in the case of QAM there's the added complication of different carrier amplitudes which I'll neglect.

Occasionally and randomly, the distribution of phases will become less-evenly-balanced and, for example, if 200 carriers take on phase 'A' at the beginning of a new symbol, 9.8% more of the carriers will have that particular phase. 'Phase A' could mean "increase from 0 + j0 to 1 + j1 in the first quarter cycle", in which case the voltages of 200 carriers would be increasing, initially, in synchronism – so their voltages would add. In time, the fact that these carriers have slightly different frequencies means their apparent synchronisation would collapse, but for the brief duration of their 'synchronism' (a fraction of the 224 μs symbol period) the instantaneous power of the COFDM signal would be increased. Since there are 1705 carriers, the power of the multi-carrier signal is, on average, about 32 dB greater than the power of an individual carrier. In the unlikely event of all carriers taking on the same phase, their voltages would all add, briefly, making the signal power about 64 dB greater than the power of an individual carrier, so the ultimate peak-to-mean ratio of the present UK DTT signal is about 32 dB.

However in practice, with the symbol duration and the strength of forward error-correction used, it is found that the peak-to-mean ratio can be clipped to 7 or 8 dB before causing appreciable loss of data integrity. If this degree of clipping is applied at source and the bandwidth of the signal is preserved, the peak-to-mean ratio should not increase in the signal path.

The same principle applies when several OFDM signals are combined (i.e. added arithmetically) and fed into a transmission line but in this case there is no clipping. If there are 6 DTT signals of equal power, the peak-to-mean ratio of the combined signal can be as much as 7.8 dB (6×) greater than that of each one, so the peak-to-mean ratio of the combined signal can be as much as 16 dB.

If the analogue television power per feeder is 25 kW (peak-sync.) per service and this is to be replaced by DTT transmissions, with no change to the transmitting antenna characteristics, each at -7 dB relative to that 25 kW then the average DTT power per transmission will be 5 kW. In view of the 8 dB peak-to-mean ratio of the signal as generated, occasional brief peaks in each DTT signal will have 8 dB greater power; that is, about 32 kW. If 6 of these signals are to be combined, a worst case peak could reach 1.19 MW whilst the average power will be 30 kW. Comparing these figures with the specifications of the example cable, there would be healthy safety factors of about 4.9 dB (or 3.1×) in respect of average power and 3.9 dB (or 2.4×) in respect of peak power – for the example cable, larger ratios than may be used at present for analogue television.

Incidentally, by 'average power' I mean that which would be measured using a slowly-acting power meter based on an RF load coupled to a temperature transducer such as a thermistor. Measuring the 'peak power' would require some kind of SHF storage-oscilloscope, or a spectrum analyser with Fourier-transform processing.

As long as the occasional peaks are brief and contribute little to the average power, there should be no need to specify the average-power rating of the cable to match the peak power, although from time-to-time suggestions have circulated in the industry that the voltage-handling capabilities of existing analogue antenna systems would only permit two conversions at -10 dB relative to the present analogue power. Perhaps this is in respect of limitations in other components like combiners, power dividers or antenna elements.

This gets rather more difficult when the plan is conversion at -4 dB meaning, for this example, an average DTT power per transmission of 10 kW and the sum of six at 60 kW. With the combined signal having 16 dB peak-to-mean ratio, the worst-case peak power would be a whopping 2.4 MW which might require the next size up, 8" feeder and a lot more expenditure (e.g. mast strengthening as well). The same manufacturer's 8" feeder has a peak-power rating of 4 MW.

It is interesting to consider how close these cables are to supporting waveguide modes. If the 6⅛" cable has $a = 33.0$ mm and $b = 77.9$ mm, and the calculated effective $\varepsilon_r = 1.063$ also applies to waveguide propagation (which is likely), then at its highest specified frequency of 860 MHz the actual wavelength in the effective dielectric is 338 mm whereas the TE_{11} cut-off wavelength is 317 mm; a ratio of only 1.07×. Putting these values into the expression given in Section 8, $\alpha = 0.0069$ neper m^{-1} corresponding to about 0.06 dBm^{-1} when considered as a power ratio. However, for a high-power television transmitting station the overall length of each feeder can be expected to be 100 m or more, and 6 dB / 100 m looks a bit more healthy! It's still not much attenuation though so, presumably, a great deal of care is needed to maintain symmetry when installing and terminating the cable to avoid the occurrence of 'trapped modes'.

If the 8" cable has a = 44.2 mm and b = 104.4 mm and the same ε_r = 1.063, then at its *lower* maximum frequency of 650 MHz (Channel 43 in Band V) the actual wavelength in the effective dielectric is 448 mm whereas the TE_{11} cut-off wavelength is 425 mm, so the ratio is an even smaller 1.05×. The largest diameter offered by the example manufacturer is 9" cable (with 5.8 MW peak-power rating), for which the maximum frequency is given as 560 MHz (Channel 32 in Band IV).

There is one further complication, concerning re-configuration of transmitting stations during maintenance work. Analogue television transmitting stations can be operated at -6 dB ERP for short periods because of the relatively 'graceful degradation' characteristic of analogue VSB-PAL, and this allows half an entire high-power station to be switched off: all the power amplifiers that feed one of the final combiners that feeds one of the feeders that feeds one half of the antenna system[†]. For DTT, reduced-power working in this way would inflict blank screens on many viewers so the ERP needs to be maintained or, at least, reduced by a smaller amount such as 3 dB. Any plan to increase the power fed to one half of the antenna system would require even greater power ratings for the feeders but, for example, the Sutton Coldfield high-power station will need to use Channel 46 for DTT (a centre-frequency of 670 MHz) so it might not be possible simply to re-equip it with 8" feeder.

The only practical solution, in some cases, means relenting about not changing the antenna characteristics and increasing the antenna gain by increasing the number of tiers of antenna elements, moderating the power rating required of the feeders and, perhaps more importantly, the amplifiers.

10.2 Cables and interfaces for instrumentation

The use of 50 Ω cables and interfaces is now more-or-less a worldwide standard for professional RF equipment and installations (apart from microwave-link and satellite-link IF sub-systems which, for some reason, are still manufactured with 75 Ω interfaces).

The fact that 50 Ω is even used for feeders at high-power transmitting stations suggests the convenience of having all parts of a complicated system working at the same impedance outweighs the 50 / 30 = 1.7× increased power-handling benefit of using 30 Ω. At least, it has so far!

Some say [4] the choice of 50 Ω is on account of ε_r = 2.26 for polythene, and 76.7 / √2.26 = 51 Ω which has been rounded down to a convenient number. So Z_0 = 50 Ω is close to the impedance for lowest loss in co-axial cables with solid polythene dielectric.

Whilst this is eminently plausible, an alternative explanation is that this value was chosen before the days of polythene, for an air-spaced line. The geometric mean of 30 Ω and 76.7 Ω is Z_0 = 48 Ω, which has been rounded up to a convenient number. This would mean the $z/30$ fractional increase of impedance in respect of the maximum power rating on account of dielectric breakdown ($P \propto 1/Z_0$) is balanced by the $76.7/z$ fractional decrease of impedance in respect of conductor loss ($\alpha \propto 1/Z_0$):

$$z/30 = 76.7/z \quad \text{so} \quad z = \sqrt{76.7 \times 30} = 48 \text{ ohms}$$

Some radically different choice such as 30 Ω would alter the ratio of ratings for maximum voltage (flash over) to maximum current (overheating) and would change the inherent peak-to-mean ratio of the cable. Although the designers of transmitting stations in the 1930s would not have had to contend with OFDM, they certainly had AM and other single-carrier modulation schemes and were undoubtedly aware of dielectric breakdown coinciding with modulation-envelope peaks. An AM modulation index of unity or 100% corresponds to a peak-to-mean envelope-voltage ratio of 2× or a peak-to-mean power ratio of 4×. This is very small in comparison with the (31×) specification of the modern-day large-diameter air-spaced cable considered in Section 10.1 but the cables available then were probably considerably more primitive. Other schemes like interrupted CW (for Morse code) could exhibit greater peak-to-mean ratios. SSB was in use for cabled telephony by then, and trans-Atlantic radio telephony experiments were reported as early as 1925 [24], but SSB may not have been applied widely to professional radio until the 1950s.

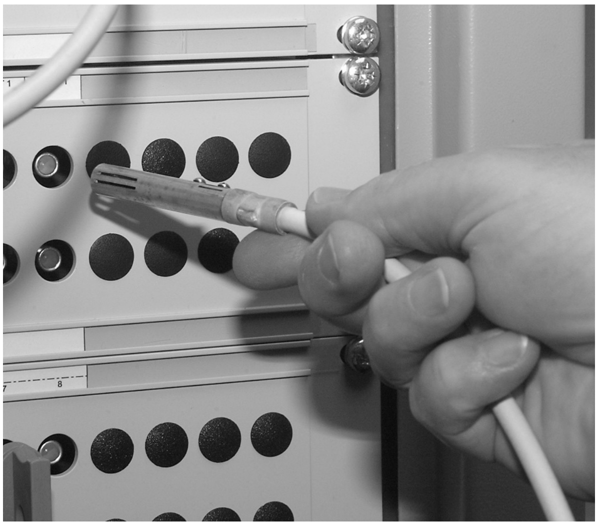
[†] Actually this is a simplification. The high-power television transmitting stations in the UK are generally arranged so that for each service the outputs of both power amplifiers are combined first. The combined signal is then split and fed to the final combiners, feeders and antenna halves. This overcomes the so-called 'Penge effect' (named after a London suburb near Crystal Palace) whereby minima in the vertical radiation pattern of the complete array are affected by differences between the power amplifiers.

However, I have been unable to find a clear account from a believable source to confirm absolutely either of these possible origins.

According to information on the worldwide-web, AT&T claim [25] to have invented co-axial cable in 1929, although Wikipedia [26] has that it was patented in 1884 by Ernst Werner von Siemens, and one of Nikola Tesla's patents from 1894 [27] clearly shows a co-axial line 'to prevent loss by dissipation or interference by induction' … 'in any system of electrical transmission or distribution in which currents of *excessively* high potential are employed, and more particularly, when the frequency is high'. AT&T made their first commercial use of co-axial cable in 1941, for long-distance telecommunications, and this was an air-spaced line with plastic beads to hold the inner conductor in place. Polythene was invented by ICI in 1931 [28] and presumably exploited commercially several years later, while polystyrene was first produced commercially in 1937. It will be seen in the next section that flexible co-axial cables were in use by 1937, but perhaps it was the large-scale production of radio and radar equipment in WWII that called for standardisation of interface impedances.

10.3 Video cables and the MUSA

Analogue and digital video cabling uses 75 Ω, generally with solid polythene dielectric. The choice of impedance in this case is evidently not for lowest loss and might be related historically to the availability of MUSA plugs and sockets, designed in the 1930s, and the ideal counterpart for patch-bays to the 'Post Office 316' jack plugs and sockets that had already been widely adopted for radio broadcasting. The PO316 (left) and MUSA (right) connectors are shown below.



The MUSA, reported in 1937 [29], was a short-wave phased array of rhombic antennas for long-range communications reception which used miles of nitrogen-filled co-axial lines made from 1" copper plumbing pipe with a ¼" copper rod inner conductor, yielding 78 Ω impedance. These led to a patch bay where different configurations could be selected using flexible co-axial jumper cables. Matching was critically important in this complicated system and there is no mention of wideband transformers in the documentation so the MUSA connectors were most likely designed for 78 Ω although I recall the BBC internal stores catalogue used to state 'Impedance: 40 to 80 Ω'! Another oddity is that in the BBC in the late 1950s, MUSA plugs were referred to by the designation 'PO No. 1'.

Where BNC (Bayonet or Baby Neill and Concelman) connectors [30] are used for analogue video connections, it has been common practice [31] to use the 50 Ω version of the connectors (with 75 Ω cables) because they were believed to be slightly more rugged. Folklore seems to abound about the diameters of the inner-conductor pins of 50 Ω and 75 Ω BNC plugs [32] but in my own experience I have always found them to be the same – it's the greater ratio of air to PTFE in the 75 Ω BNC socket that gives it its higher characteristic impedance. On the other hand, the pins of N and C plugs for these two impedances have greatly differing diameters.

Because an analogue video signal occupies a huge fractional bandwidth (25 Hz to 5 MHz covers the best part of 18 octaves), it is inevitable that co-axial lines any longer than a few metres require equalisation. Over the history of analogue television the BBC has equipped itself with equaliser designs for all conceivable applications. The frequency dependence of the loss is predominantly on account of the skin effect, for which $\alpha \propto \sqrt{f}$, but this loss should tend towards a constant value (per metre) at very low frequencies as the skin depth exceeds the actual thickness of the conductors. For example, if the outer conductor were 0.5 mm thick copper the skin depth would reach this thickness at $f = S / \pi\mu_0 d^2$ = 17.8 kHz. This appears not to have been taken into account explicitly, but the loss at such low frequencies is most likely relatively very small.

General group-delay equalisation seems not to have been an acute problem in video cabling but many of the BBC equalisers had delay correction at the 4.43 MHz colour sub-carrier frequency where delay errors could have a disproportionate effect on the PAL system.

If the choice had been made many years ago to use 50 Ω instead, perhaps if someone had conceived a 50 Ω co-axial 'jack plug' and socket for another application, it probably wouldn't have changed the course of television history. It would have increased the output-current demand on amplifiers to drive co-axial lines, but it would have reduced the gain needed to recover signals from long lines. The conductor losses of 75 Ω and 50 Ω polythene-dielectric cables have a ratio of 1.12:1. The most common cable used in the BBC for video installations is known as PSF1 / 3 [33] and this 75 Ω cable has a loss of 3.3 dB / 100 m at 10 MHz. A 50 Ω counterpart with the same outer-conductor diameter would exhibit a loss of about 2.9 dB / 100 m.

The choice of 75 Ω, as opposed to 50 Ω, for digital television interfaces (e.g. SDI) probably has little to do with physics and more to do with the ubiquity of 75 Ω hardware in the television industry.

10.4 Domestic television antenna installations

75 Ω polythene-foam-dielectric cable is common in the UK for terrestrial television antenna installations. The foam dielectric provides characteristics similar to air, and the chosen impedance seems a reasonable compromise between 76.7 Ω for minimum loss and matching the 73 Ω terminal resistance of a λ/2 dipole. Although Yagi-Uda antennas are used, almost without exception, nowadays, and this type of antenna can have any terminal impedance the designer chooses (or can manage!), 75 Ω has become the *de facto* standard. This may be traceable to the widespread use of dipoles and simpler arrays for the original major UK roll-out of television in Band I in the early 1950s. We even have the bespoke 'Belling Lee' plugs and sockets which are sometimes advertised as having a nominal impedance of 75 Ω although, unlike professional connectors like the BNC, they are not designed for uniform characteristic impedance – they're just small enough to do little damage in this less-critical application.

Co-axial cable sold for domestic installations is designed to be relatively inexpensive and was often of poor quality with outer conductors made of loosely-woven wire strands, introducing additional loss on account of radiation and allowing ingress of interference. However, the arrival of satellite broadcasting with L-Band interfaces between front-end and set-top box has necessarily made rather better cables available at domestic prices.

Incidentally, at frequencies for which dielectric loss is smaller than conductor loss, the purpose of using foamed rather than solid polythene would appear to be to reduce the ε_r so as to increase the required diameter of the inner conductor in order to reduce its conductor loss – rather like the "vacuum cleaners don't suck" argument!

This is readily demonstrated using the expressions in Section 7.3 to compare the attenuation values for two 6 mm diameter cables at 600 MHz, one having solid polythene dielectric and the other air dielectric, and appropriate values of ψ to yield 75 Ω in each case (6.55 and 3.49, respectively).

$$\alpha = 0.0172 + 0.00293 \text{ (neper m}^{-1}\text{) for solid polythene}$$

$$\alpha = 0.0102 + 0 \qquad \text{(neper m}^{-1}\text{) for air}$$

… where the first and second terms in each case represent conductor loss and dielectric loss. The reduction of conductor loss is more than twice the reduction of dielectric loss.

10.5 Computer networks

Nowadays, most computer data networks use cables containing multiple un-screened (or ‘unshielded’ in the USA) twisted pairs, but co-axial cables were used in the early days. Although 50 Ω cable and interfaces were specified for the original 10 Mbs^{-1} IEEE 802.3 Ethernet (also known as 10Base5), for one particular system ‘ARCnet’ (Attached Resource Computer Network [34]) the specified cable impedance was 93 Ω and a special RG-62A / U cable was manufactured for this application. This scheme offered operating speeds approaching 20 Mbs^{-1}. I wonder what led the designers to choose that impedance.

Although I’ve based the whole treatment in this document on excitation by a sinusoidally alternating source, it is of course possible to derive the whole lot by time-domain methods which would be more-directly applicable to signals in computer networks like this.

11. Conclusions

I have presented here some facts you may or may not have known about co-axial cables. I have derived from fairly basic principles (not first principles but, perhaps, 1½th principles!) expressions for the minimum field strength, characteristic impedance, maximum power handling on account of dielectric breakdown, conductor loss and dielectric loss, and I have identified optimum values where they exist. Where I haven’t tackled the full derivation, for instance in the case of waveguide propagation, I have offered some appropriate references.

I expect I have uncovered the most likely physical reasons for the ubiquitous choices of 50 Ω and 75 Ω characteristic impedance, although the historical truth about the former still seems elusive. I haven’t dealt in detail with dissipation of heat in connection with loss, or the fact that stranded wires are used for the conductors in most flexible co-axial cables, but I have demonstrated that co-ax has an inherent peak-to-mean ratio and I’ve illustrated what this might mean in the context of television switch over.

I hope this White Paper proves useful to one or two.

I wish to thank all those colleagues who have helped me in this endeavour by proof reading and checking the arithmetic, particularly Dave Darlington, Jonathan Stott, Peter Shelswell and Nick Wells.

12. References

1. HARVEY, A.F. Microwave engineering. London New-York Academic Press, 1963. Page 11
2. KRAUS, J.D. Electromagnetics. Fourth edition. McGraw-Hill, Inc., 1992, ISBN 0-07-035621-1. Page 589, Q12-2 'Why are the most common coaxial line impedances 50 and 75 Ω?' with the answer on Page 837 given as [3]
3. Microwave Journal, State of Art Reference (supplement, not part of the Euro-Global edition), September 1990, Page 42
4. An article in 'EDN' magazine, available in September 2005 from http://www.edn.com/contents/images/47191.pdf and the follow-up article http://www.edn.com/contents/images/60931.pdf
5. KRAUS. *Ibid.* Page 65
6. *Ibid.* Pages 475 to 499
7. *Ibid.* Pages 236 and 237
8. ROSA, E.B. The self and mutual inductances of linear conductors. *Bulletin of the Bureau of Standards*, Vol. 4, No. 2, 1908. Page 303
9. An explanation of binomial series, available in September 2005 from: http://mathworld.wolfram.com/BinomialSeries.html where this is credited to Eric W. Weisstein
10. KRAUS. *Ibid.* Page 492
11. HEAVISIDE, O. Electromagnetic theory. Volume 1. Benn Brothers Ltd., London, 1922. Pages 366 to 369 (the material in the three volumes of this book was previously serialised in *The Electrician* between 1885 and 1887)
12. JORDAN, E.C. and BALMAIN, K.G. Electromagnetic waves and radiating systems. Second edition. Prentice-Hall, Inc., 1968. Page 158
13. KRAUS. *Ibid.* Page 551
14. An explanation of the 'Lambert W-function', available in September 2005 from: http://mathworld.wolfram.com/LambertW-Function.html, again credited to Eric W. Weisstein
15. 'Technical Description RF Cables', available in May 2006 from: http://www.rfsworld.com/userfiles/pdf/technical-description-rfcables.pdf
16. KRAUS. *Ibid.* Page 442
17. Times Microwave catalogue of co-axial cables, available in September 2005 from: http://www.timesmicrowave.com/content/pdf/tl14/tl14_catalog.pdf
18. US military specifications for co-axial cables, MIL-C-17, available in September 2005 from: http://www.dscc.dla.mil/Programs/MilSpec/DocSearch.asp
19. JORDAN and BALMAIN. *Ibid.* Pages 726 to 729
20. SCHELKUNOFF, S.A. Electromagnetic waves. Van Nostrand, Inc., 1943. Page 196
21. KRAUS. *Ibid.* 671
22. HARVEY. *Ibid.* Page 20
23. Data on 6⅛" air-dielectric co-axial cable manufactured by RFS, available in September 2005 from: http://www2.rfsworld.com/RFS_Edition3/pdfs/HELIFLEX_Air_Cable_34-47.pdf
24. Heising, R.A. Production of single sideband for trans-atlantic radio telephony. Proc. IRE, Vol. 13, No. 3, June 1925. Pages 291 to 312
25. Notes on the history of AT&T 'long lines' unit, available in September 2005 from: http://www.bellsystemmemorial.com/longlines_book1.html, http://www.att.com/history/milestones.html and, for a photograph of a piece of their first cable: http://www.att.com/history/milestone_1941.html

26. The entry in Wikipedia (an online encyclopedia) on co-axial cables, available in September 2005 from: http://en.wikipedia.org/wiki/Coaxial_cable

27. A US patent specification filed by Nikola Tesla, available in September 2005 from: http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=/netahtml/srchnum.htm&r=1&f=G&l=50&s1=514167.WKU.&OS=PN/514167&RS=PN/514167

28. A brief history of the development and application of plastics, available in September 2005 from: http://www.tangram.co.uk/TI-Polymer-Plastics_Timeline.html

29. FRIIS, H.T. and FELDMAN, C.B. A Multiple Unit Steerable Antenna for short-wave reception. Proc. IRE, Vol. 25, No. 7, July 1937. Pages 841 to 917

30. The history of the BNC connector according to one of its inventors, available in September 2005 from: http://www.amphenolrf.com/products/bnc.asp

31. A description of the differences between 50 Ω and 75 Ω BNC connectors, available in September 2005 from: http://sosnick.uchicago.edu/BNC_50_75.html

32. An account of some of the issues, and some of the 'folklore' associated with 50 Ω and 75 Ω BNC connectors, available in September 2005 from: http://www.canford.co.uk/techzone/BNCConnectors.PDF

33. A specification for the BBC-standard PSF1/3M co-axial cable for video, available in September 2005 from: http://www.canford.co.uk/commerce/resources/catdetails/2207.pdf

34. A description of the ARCnet standard for computer networks, available in September 2005 from: http://www.comptechdoc.org/independent/networking/guide/netarcnet.html