The Bruene Directional Coupler and Transmission Lines

Gary Bold, ZL1AN. morseman@nzart.org.nz

Abstract

The Bruene directional coupler indicates "forward" and "reflected" power. Its operation is invariably explained using transmission line concepts, which leads some to wrongly believe that it must always be connected *directly to* a line. There are also many misconceptions about what happens *on* a transmission line, and in particular, what happens at the source end. This document addresses these misconceptions.

Unfortunately, many derivations require complex exponential notation, phasor representations of voltage and current, and network theorems. There isn't any alternative if rigor is to be maintained.

Section 1 gives some background on the Bruene coupler.

Section 2 explains its operation when terminated in a resistance.

Section 3 works a simple numerical example on a matched transmission line and load.

Section 3 repeats this calculation for a *mismatched* line and load.

Section 4 explains the correct interpretation of the powers read by the Bruene meter.

Section 5 explains that a mismatched source *always* generates additional reflections, and shows how the steady-state waves on the line build up by summing these reflections.

Section 6 develops the theory of section 6 further, and shows how the final waves build up in the mismatched line example given in section 4.

Four appendices follow:

Appendix A develops the general theory of line reflection where both source and load are mismatched, and the line may be lossy.

Appendix B works a simple example to illustrate appendix A.

Appendix C explains what happens upon reflection from a Thevenin source.

Appendix D shows, using the Poynting vector, that power *always* travels from the source to the load at all times, everywhere on the line.

1 The Bruene Coupler

Warren Bruene, then W0TTK, a long-time *Collins* design engineer, published the original article describing the directional coupler that bears his name in QST in April, 1959.[1] This was used in the *Collins* 302C Wattmeter in 1960.

This is *not* the only form of directional coupler used in SWR/power meters. Strip-line and resistance-bridge detectors are also common, though less so now, due to the ease with which suitable small toroids can be obtained, and because the Bruene coupler can be permanently inserted in the line, and is not frequency dependent.

The operation of the Bruene coupler is traditionally derived as if it's inserted *in* a transmission line, using the concepts of "forward" and "reflected" waves, which are assumed to exist on the line *before* and *after* it, and which flow *through* it. But the coupler also works when connected to the *input* of a line, where there is no line on the "input" side, or when connected to a transmatch, or even to a pure resistance - which must always be done to calibrate it. There are no standing waves inside a transmatch or resistance, and the coupler itself doesn't contain a transmission line, so there must be an alternative way of explaining its operation. This we now do.

[1] Download Bruene's QST article from http://www.arrl.org/tis/info/pdf/5904024.pdf

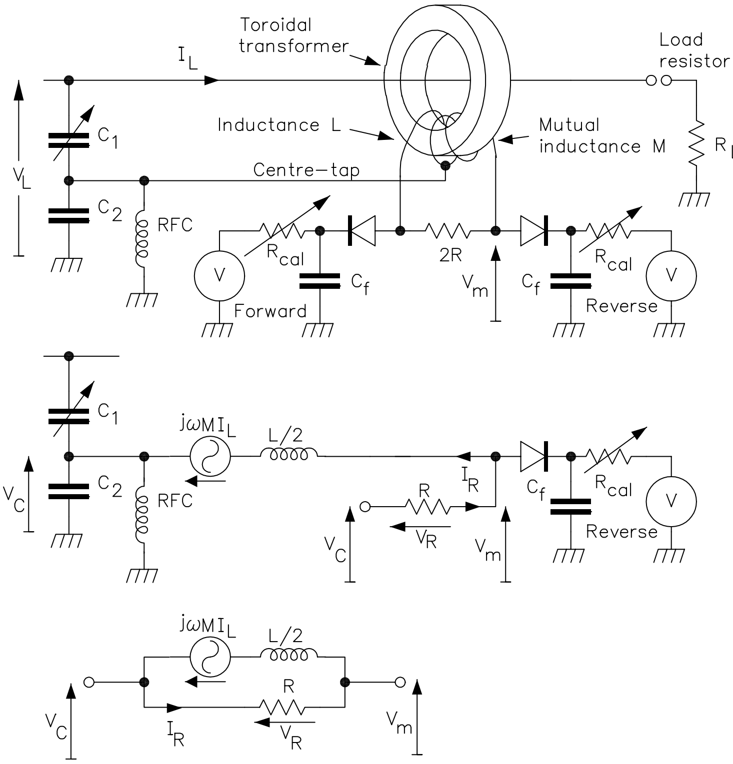


Figure 1: Top: Practical Bruene directional coupler. Middle: Simplified to aid in derivation of power estimation equations. Bottom: Even further simplified.

2 The "Practical" Bruene Meter

If you look inside a typical modern meter using the Bruene principle, you probably won't see the circuit almost always used to explain its operation (Bruene's original version), which was published in *Break-In* September/October 2006,[2] you'll see a version of the more practical circuit of figure 1, top.

Figure 1 differs in that there's only *one* voltage-sampling capacitive divider, which is connected to the *mid-point* of the secondary of the current-sampling toroidal transformer. The "forward" and "reverse" (diode and smoothing capacitor samplers) are *not* connected in bridge configurations, but to voltage summer circuits. V_L and I_L are the instantaneous (AC, signal) voltage and current on the sampling wire that runs from the input to the output, through the toroid. This wire is *not* a transmission line. The output is connected to a pure resistance R_L. The inductor labelled "RFC" holds the relatively high impedance midpoint of the capacitive divider at DC ground, necessary because the diode rectifying circuits are also referenced to ground. In some meters it is omitted. Often it's parallelled by a resistor which doesn't affect the derivation below.

Figure 1, middle, shows that portion of the circuit involved in deriving the "reverse" voltage. The transformer secondary has been replaced by its equivalent circuit, a voltage generator proportional to, and in quadrature with, the line current, in series with half of the secondary inductance.

Figure 1, bottom, shows an even further simplification, retaining only the elements necessary to derive the equations.

Neglecting the loading effect of the diode detection circuit (the calibration resistance R_{cal} is much higher than the $2R$ connected across the secondary), we see that the midpoint of the resistance $2R$ must be, by symmetry, at the same voltage as that of the secondary centre-tap, set by the capacitive voltage divider. Hence we can isolate it from the "forward" components by substituting *half* of the terminating resistance, R, terminated by the mid-point voltage V_c. Then the current through R is given by

[2] "*SWR Meters; How they work*" David Conn, VE3KL, *Break-In*, September/October 2006, pp 8 - 10. Reprinted from *The Canadian Amateur*, July/August 2005.

$$\begin{aligned}
I_R &= \frac{j\omega M I_L}{R + j\omega L/2} && (1)\\
\text{but } \omega L &\gg R \qquad \text{(we choose this)} && (2)\\
\text{so } I_R &\approx \frac{2M}{L} I_L && (3)\\
V_R &= R I_R = R\frac{2M}{L} I_L && (4)\\
\text{but } I_L &= \frac{V_L}{R_L} \quad (V_L, I_L \text{ are related by the load resistance, } R_L) && (5)\\
\text{therefore } V_R &= R\frac{2M}{L}\frac{V_L}{R_L} = V_L \frac{2M}{L}\frac{R}{R_L} && (6)\\
\text{using Kirchhoff's voltage law, } V_m &= V_C - V_R && (7)\\
&= V_L\left(\frac{C_1}{C_1 + C_2} - \frac{2M}{L}\frac{R}{R_L}\right) && (8)\\
\text{but } C_2 \gg C_1 & && (9)\\
\text{so } V_m &\approx V_L\left(\frac{C_1}{C_2} - \frac{2M}{L}.\frac{R}{R_L}\right) && (10)
\end{aligned}$$

The AC voltage V_m is rectified, smoothed and indicated by the voltmeter. To calibrate, we make $R_L = 50\,\Omega$ and adjust C_1 to make the two terms inside the bracket equal, so that $V_m = 0$. Physically, this means that the voltage derived from the current-sampling toroid is exactly equal and opposite to that derived from the voltage-sampling capacitive divider, so that their algebraic sum is zero.

The corresponding voltage at the detection point of the "forward" circuit is given by the same equation, *except* that the terms inside the brackets *add* instead of *subtracting*, because the voltage induced by its half the current transformer is in the *opposite phase*. That is,

$$\text{Forward voltage is } V_m \approx V_L\left(\frac{C_1}{C_2} + \frac{2M}{L}.\frac{R}{R_L}\right) \qquad (11)$$

If the terminating resistance is now changed in value, keeping the line voltage V_L constant, the line current, I_L will change, and V_m will rise. You can see this from the second term inside the brackets of equation 11, which is inversely proportional to R_L. Similarly, substituting any reactive impedance will cause unbalance because the current-derived waveform phase will change from 180^o. Thus, all that "zero reflected power" tells you is that the impedance seen at the meter output is exactly $50\,\Omega$, resistive. And this is true whether you're connected to a transmatch, a transmission line, or a pound of butter.

If the meter is connected directly to a transmission line, it still just sees an impedance, the input impedance of the line. But this impedance has a value which is very restricted. It's somewhere on a constant SWR circle drawn about the centre of a Smith chart normalized for the line characteristic impedance, which is the resistance which has been used to calibrate for zero reflected power. The "forward" and "reflected" readings don't allow you to determine this input impedance value, but they *do* allow you to determine *which* SWR circle they're on, and thus the SWR.

Both the "propagating wave" and "impedance terminated" treatments of the coupler given above are *equally valid*.

It's interesting that (as Bruene points out in his original article) the detected voltage *decreases* as the number of turns on the toroid secondary *increases*. This is because a toroid's mutual inductance, M is proportional to the number of turns, n, while inductance, L, increases as n^2. However, you can't reduce n arbitrarily to achieve greater sensitivity, because this progressively limits the lowest operating frequency, which is proportional to L.

Now consider what happens on a transmission line itself, in matched and unmatched situations. For simplicity, we invoke a lossless line of length $\lambda/4$ having characteristic impedance of $50\,\Omega$, and purely resistive loads. For a lossless line, z_0 is also resistive, and is sometimes called R_o.[3]

[3] The results are valid for lossy and reactive loads as well, but the algebra becomes tedious.

3 Example 1: A Matched Line

Figure 2 (a) shows a lossless line having characteristics impedance $z_o = R_o = 50\,\Omega$.[4] It's terminated in a pure resistance R_L also of $50\,\Omega$, and is fed from a source that develops a voltage $V_{in} = 5V_{rms}$ at its input.

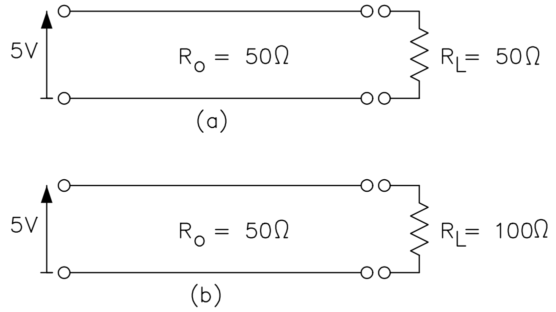


Figure 2: (a) A lossless line

The input impedance of the line is resistive, and the same as R_o, or $50\,\Omega$. The input power, and the output power passed to the load, P_{load}, are the same, and are given by

$$P_{in} = P_{load} = \frac{V_{in}^2}{R_o} = \frac{5^2}{50} = 0.5 \text{ Watt} \tag{12}$$

In conventional terminology, there is no "reflected wave". A single "forward wave" propagates down the line and its power is completely absorbed by the load. The amplitude of this wave is equal to the input voltage, 5 Volt.

4 Example 2: A Mis-matched Line

Figure 2(b) shows the same line, now terminated in a pure resistance of $100\,\Omega$. The line is now mismatched, and acts as a quarter-wave impedance transformer. Such a line's impedances obey the equation

$$R_{in}R_L = R_0^2 \tag{13}$$

$$\text{so } R_{in} = \frac{R_o^2}{R_L} = \frac{50^2}{100} = 25\,\Omega. \tag{14}$$

and the input power developed in this resistance by the same input voltage will be

$$P_{in} = \frac{5^2}{25} = 1 \text{ Watt} \tag{15}$$

Again, all of this power is passed to, and is dissipated in the load.[5] We can calculate the voltage across the load from the power dissipated in it, using conservation of energy.

$$P_{load} = P_{in} = \frac{V_L^2}{R_L} \tag{16}$$

$$\text{so } V_L = \sqrt{P_{in}R_L} = \sqrt{1 \times 100} = 10 \text{ volt.} \tag{17}$$

[4] The characteristic impedance of a lossless line is always real, purely resistive.

[5] The power delivered to the load is *higher*, even though the line is mismatched in the second example, because we are applying the *same* voltage across a *lower* input resistance.

Because the line is now not terminated in its characteristic resistance, it will now have a standing wave ratio, SWR, which we calculate using the voltage reflection coefficient, ρ, given by[6]

$$\rho = \frac{R_L - R_o}{R_o + R_L} = \frac{50}{150} = \frac{1}{3} \tag{18}$$

$$\text{and } SWR = \frac{1+\rho}{1-\rho} = \frac{4}{2} = 2. \tag{19}$$

In conventional (and correct) treatments, the resulting voltage variation on the line is now interpreted as the sum of a *forward* wave, of amplitude V_f, and a *reverse* wave, of amplitude V_r. These propagate independently, and each behaves as if travelling in a *matched* line of characteristic resistance R_o, which makes them both simple and useful. They are related by $V_r = \rho V_f$. The maximum and minimum voltages on the line are given by the constructive and destructive interference of these waves, so that

$$V_{max} = V_f + V_r = V_f\,(1+\rho) \tag{20}$$

$$V_{min} = V_f - V_r = V_f\,(1-\rho) \tag{21}$$

In this special case of resistive termination where $R_L > R_o$, the load voltage will be at a SWR *maximum*, and the voltage a quarter wavelength back, at the input, will be at a *minimum*. We can use these values to find V_f and V_r. Adding the equations above,

$$V_{max} + V_{min} = 2V_f \tag{22}$$

$$\text{so } V_f = \frac{V_{max} + V_{min}}{2} = \frac{10+5}{2} = 7.5 \text{ volt} \tag{23}$$

$$\text{and } V_r = \rho V_f = \frac{7.5}{3} = 2.5 \text{ volt.} \tag{24}$$

Before considering the significance of these results, we calculate the powers that each wave apparently carries.

5 Power Calculations on the Mismatched Line

Directional coupled SWR meters are usually calibrated in terms of "forward" and "reflected" *power*, though they do not measure power directly. They detect the *voltages* and *currents* of these two waves, and deduce the powers indicated by assuming that each propagates in a line of characteristic resistance R_o, the same as that of the calibration, "zero reflected power" resistance. This is why the scales are nonlinear, and roughly follow a square-root law, modified by the practical conduction characteristics of the rectifying diodes used in the AC (signal) to DC rectifiers.

The power in the forward and reflected waves, P_f and P_r respectively, are given by

$$P_f = \frac{V_f^2}{R_0} = \frac{7.5^2}{50} = 1.125 \text{ Watt} \tag{25}$$

$$P_r = \frac{V_r^2}{R_0} = \frac{2.5^2}{50} = 0.125 \text{ Watt} \tag{26}$$

$$\tag{27}$$

[6]The derivation of ρ, also called ρ_v is done in all standard texts, including mine, *"Electromagnetism" Gary E.J. Bold,* Physics Department, University of Auckland, latest edition, \$15 (NZ), chapter 11. This text also gives a simplified, traditional treatment of the Bruene coupler. To be rigorous, we should use the *magnitude* of ρ, or $|\rho|$, but here ρ is real and positive, so I've left the magnitude signs out.

The *difference* of these powers is 1 Watt, the same as the power that we have independently shown to be dissipated in the load, using conservation of energy. This can be shown to be always true, for *any* line length and terminating impedance, although the algebra becomes tiresome for the general case. This justifies the conventional rule

> "To find the true power passed to the load, *subtract* the "reverse" power from the "forward" power.

The "forward" power would be better called "total" or "forward plus reflected" power, and some authors have advocated this. The *only* reason that directional couplers are calibrated in power for both "waves" is to allow this subtraction to be done.

This relationship holds even in extreme practical cases, where a 100 Watt transmitter operating into a very mismatched line may indicate 900 Watt forward power and 800 Watt reflected power. *Nothing* in the system is generating 800 Watt. It is only the *difference* in the powers that has any significance.

6 Formation of "Forward" and "Reflected" Waves

It's often assumed that the "forward" wave is directly generated by the voltage applied to the line, and the "reflected" wave, with amplitude related to the load impedance is derived from this, and "carries power back" towards the source.

However, in example 2 the amplitude of the forward wave, V_f, is constant at all points on the line, and has been shown to be 7.5 volt. This is *greater* than the voltage applied to the line input! Therefore this wave *cannot* have been excited *directly* by the input voltage alone. In fact, *both* of the forward and reflected waves are superpositions of (in principle) an infinite number of counter-travelling waves. The proof of this was commonly performed in formal University courses 50 years ago[7] and still exists in some texts.

We first consider a simplified situation, where the source voltage is derived from a zero-impedance source, so that V_1 remains unaltered by any change in the line input impedance. The general development of the relationships derived here is given in appendix A of this document.[8]

Consider a voltage V_1 suddenly applied at one end of a (possibly lossy) transmission line, causing a wave which travels towards the load. Since it doesn't know about the load until it gets there, it travels through the appropriate characteristic impedance of the line. Let its complex amplitude when it reaches the load be V_f, where[9]

$$\begin{aligned}
V_f &= V_1 e^{-\gamma\ell} & (28)\\
\text{where } \gamma &= \text{the complex propagation constant of the line,} & (29)\\
\gamma &= (R + j\omega L)(G + j\omega C) & (30)\\
\text{where } R, L, C \text{ and } G &= \text{the standard line constants, "quantity per unit length"} & (31)\\
\ell &= \text{the physical line length.} & (32)
\end{aligned}$$

Satisfying boundary conditions at the load z_L means that a reflected wave is now generated, with complex amplitude V_r at the load, where

[7] I was shown this proof, when a student, in *Radiophysics III* in 1960 by Professor Kurt Kreilesheimer.

[8] the general proof is performed in some texts, see for example *Theory and Problems of Transmission Lines*, Robert Chipman, Schaum's outline series, McGraw-Hill, 1968, chapter 8. It applies only to a *linear* source, but we'll consider the implications of this when we come to it.

[9] As is usual in these calculations, we represent wave amplitude and phase changes by a complex exponential representing a rotating phasor in the complex plane, which enormously simplifies the algebra. "Complex" algebra does not mean "complicated", but the representation of a network quantity by "real" and "imaginary" (or "quadrature") components, carrying information about *both* amplitude and phase. The "actual" wave amplitude and phase can subsequently be found by just taking the real part of the resulting expressions. This is explained in all more advanced texts. See for example "*Linear Steady-state Network Theory*" Gary E.J. Bold, Physics Department, University of Auckland, latest edition, chapters 2 and 3, $15 (NZ)

$$\begin{aligned} V_r &= \rho_2 V_1 e^{-\gamma\ell} = \rho_2 V_f && (33)\\ \text{and } \rho_2 &= \text{ the voltage reflection coefficient at the load,} && (34)\\ &= \frac{z_L - z_o}{z_L + z_o} && (35)\\ V_f &= V_1 e^{-\gamma\ell} = \text{ the amplitude of the reflected wave arriving at the load.} && (36) \end{aligned}$$

This is the standard treatment, and generally the mathematics stops here, because ρ_2 is found to *completely* determine the SWR and impedance at any point on the line. What happens now to the reflected wave when it eventually reaches the *source* end can be, and usually is, ignored.

But electromagnetic wave theory, derived from Maxwell's equations[10] says unambiguously that there *must* be a reflection *whenever* any impedance discontinuity is encountered, and such a discontinuity must occur when the reflected wave again reaches the source, *unless* the input is fed from a linear source having an output impedance $z_s = z_o$, exactly the same as the characteristic impedance of the line. This would be a "complex-conjugate match", and cannot in general hold, since all transmitters work with all lines, and all are unlikely to have exactly this impedance. More about this later.

The reality of source reflections are well known to engineers who send short pulses down lines. This is why professional signal generators have an output impedance of $50\,\Omega$, resistive, to match the characteristic resistance of the line between source and load, removing the possibility of reflection here. The reality of such reflections is easily demonstrated using an oscilloscope and pulse-generator.

If the source is mismatched, the *second* reflection now occurring on this *first* reflected wave at the line input can also be characterized by a reflection coefficient ρ_1, and *another forward* travelling wave will be generated, of complex amplitude V_1', where

$$\begin{aligned} V_1' &= \rho_1\rho_2 V_f e^{-\gamma\ell} && (37)\\ \text{where } \rho_1 &= \frac{z_s - z_0}{z_s + z_0} && (38)\\ \text{and } z_s &= \text{ some, as yet undefined, source impedance.} && (39) \end{aligned}$$

Now there are *two* forward waves. When this one reaches the load, its amplitude will be

$$V_f' = \rho_1\rho_2 V_f e^{-2\gamma\ell} \qquad (40)$$

this second *forward* wave generates a second *reflected* wave, having amplitude

$$V_r' = \rho_1\rho_2^2 V_f e^{-2\gamma\ell} \qquad (41)$$

This process continues indefinitely. Summing, the total forward and reflected waves at the load end will be

$$\begin{aligned} (V_f)_t &= V_f\left(1 + \rho_1\rho_2 e^{-2\gamma\ell} + \rho_1^2\rho_2^2 e^{-4\gamma\ell} + \ldots\right) && (42)\\ (V_r)_t &= V_f\rho_2\left(1 + \rho_1\rho_2 e^{-2\gamma\ell} + \rho_1^2\rho_2^2 e^{-4\gamma\ell} + \ldots\right) && (43)\\ \text{dividing, } \frac{(V_r)_t}{(V_f)_t} &= \rho_2 && (44) \end{aligned}$$

Equation 44 is identical to that given by considering *only* the first forward and reflected waves, and again depends *only* on the load.

We see that the value of ρ_1 *doesn't matter*, because all terms containing it cancel. This is fortunate, because it isn't clear to me what it is. In a linear system, z_s in equation 38 would be the source impedance.

[10] These 4 vector equations were derived by Heaviside from Maxwell's original ones, which were incomprehensible to almost everybody. They are derived in all advanced electromagnetic texts, including mine. They explain *all* electromagnetic phenomena. There is no doubt about their validity.

But the *hf* transmitter output stage driving the line contains transistors operating in a non-linear manner, usually class B or C, and linear circuit theory, in particular Thevenin's theorem, isn't applicable. It may even be that the transmitter's output impedance varies over an operating cycle, but even *this* will not matter, because the same value of ρ_1 occurs in matching terms in both forward and reflected waves, evaluated at the same time, which are divided to form equation 44.

The only observable effect of this summation of forward and reflected waves is to change the input impedance of the line. For a lossless line, this is given by the (standard) equation[11]

$$z_{in} = R_o \left[\frac{jR_o \sin(\beta\ell) + z_L \cos(\beta\ell)}{jz_L \sin(\beta\ell) + R_o \cos(\beta\ell)} \right] \tag{45}$$

$$\begin{aligned}
\text{where } R_o &= z_o = \text{ line characteristic impedance,} && (46)\\
z_L &= \text{ the terminating impedance, possibly complex,} && (47)\\
\ell &= \text{ physical length of the line,} && (48)\\
\beta &= \text{ phase constant } = \sqrt{LC} = \frac{\omega}{c}, && (49)\\
\omega &= 2\pi f = \text{ angular frequency of the source,} && (50)\\
c &= \text{ phase velocity on the line.} && (51)
\end{aligned}$$

We also see that the *total* forward wave has an amplitude which is the sum of an assemblage of sub-waves, and is *not* in general equal to the applied voltage alone. This is shown by the mismatched-line example 2 above, where the line is energized by a source of 5V, and the amplitude of the forward wave is 7.5V. This amplitude is a result of the summation shown in equation 42. Similarly for the reflected wave amplitude.

We were only able to compute the actual magnitudes of the forward and reflected waves in this (carefully chosen) example because the positions of the maximum and minimum (load and source) of the resulting total wave on the line were known.

Furthermore, we see, and will next show, that these amplitudes are the *total* amplitudes under steady-state conditions, that is, after the initial "wave build-ups" have reached their asymptotic values. Actually, these "asymptotic values" are never, strictly attained, because even though both ρ_2 and ρ_1 are always less than one, so that each subsequent forward and backward wave becomes progressively smaller, neither wave component ever becomes *zero* for a lossless line.[12] However, after a sufficient number of reflections, each will become vanishingly small. For a "lossy" line, and all lines are actually "lossy", *some* power *will* be dissipated in the line.

This derivation shows that the "standard" forward and reflected waves take a finite, though small, time to form, since energy has to travel up and down the line to create the wave assemblages that are summed. However, this process can be pretty well considered "instantaneous" at hf, since typically a maximum of 10 or so line-lengths are travelled before contributions become vanishingly small - about half a microsecond on a typical 10 metre length of coax having a velocity factor of 66%.

7 Formation of Waves in Example 2

We now show that in the mismatched example considered, the steady-state, final amplitudes of the forward and reflected waves deduced agree with the summation of the multiply reflected waves that progressively build up as shown in the previous section.

For simplicity, again assume that the line is energized by a 5V source, having *zero* internal impedance z_s - that is, reflections cannot alter this input voltage. This is a special case of a time-independent, linear source, chosen to make the calculation simpler, and to show that the results hold even in this extreme case (complete reflection at the source). The reflection factor at the line input, ρ_1, defined just as that at the load end, ρ_2, is

[11] derived, again, in all standard texts.

[12] This is analogous to the exponentially changing voltage across a charging capacitor, which never actually reaches its final value either. However, we normally assume that it has done so after about 5 time constants have elapsed.

$$\rho_1 = \frac{z_s - z_o}{z_s + z_o} = \frac{0-50}{0+50} = -1 \tag{52}$$

$$\text{while } \rho_2 = \frac{1}{3} \text{ as before.} \tag{53}$$

This value of ρ_1 means that waves arriving back at the source will be *completely* reflected, but undergo a *phase change* of 180^o. The first forward wave will start with an amplitude of 5V, the same as the source. The first reflected wave resulting from this will have an amplitude of 5/3V, and travel back towards the source. At the source, it is completely reflected, but with a phase change of 180^o. However, since it has travelled a total distance of $\lambda/2$ (a quarter wave down and a quarter wave back) it will arrive 180^o *retarded* in phase. This phase retardation *cancels* with the phase change at source reflection, so it's reflected amplitude is *in phase* with that of the initial wave. The sum of these two waves will therefore be

$$V_{f1} + V_{f2} = 5 + \frac{5}{3} \text{ Volt} \tag{54}$$

This second forward wave also undergoes a load reflection, then a source reflection. It's contribution will be V_{f3} where

$$V_{f3} = \frac{1}{3} V_{f2} = \frac{5}{9} \text{ Volt} \tag{55}$$

This series of reflections continues indefinitely, until, in practice, the contribution of further reflections becomes vanishingly small. Summing, the total forward wave amplitude will be

$$(V_f)_t = V_{f1} + V_{f2} + V_{f3} + \ldots \tag{56}$$

$$= 5 + \frac{5}{3} + \frac{5}{9} + \frac{5}{27} + \ldots \tag{57}$$

$$= 5\left(1 + \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \ldots\right) \tag{58}$$

The series inside the brackets can be summed to infinity using the rule for geometric progressions,[13] giving

$$(V_f)_t = 5\left(\frac{1}{1-1/3}\right) = 5\left(\frac{1}{2/3}\right) = 5\left(\frac{3}{2}\right) \tag{59}$$

$$(V_f)_t = 7.5 \text{ Volt} \tag{60}$$

which is the *same* as that deduced earlier from a calculation involving the SWR and maximum and minimum voltage values. A similar calculation shows that the amplitude of the reflected wave is also correct, 2.5 Volt. This confirms that

- the forward and reflected wave amplitudes deduced from the SWR are indeed the steady-state, asymptotic values,
- these arise quite naturally from assuming that the amplitude of the *first* forward wave is the same as that of the source voltage, and adding *further reflections from both the source and load.*

Appendix A shows that such a summation also gives the correctg steady-state wave amplitudes when the source impedance, z_s has a finite value, that is, for any general linear Thevenin source. In this case the open-circuit voltage of the source is reduced to a lower value at the line input because of the line's loading

[13] If a series can be written as $S = 1 + r + r^2 + \ldots$ where $r < 1$, the sum to n terms is $S = \frac{1-r^n}{1-r}$. As $n \to \infty$, the second term in the numerator, r^n, becomes vanishingly small, and the sum becomes $S = \frac{1}{1-r}$.

effect on the source. The derivation also holds for a lossy line of any length, terminated in any impedance. However, in these general cases the algebra becomes unwieldy since ρ_1 and ρ_2 are in general complex, so that subsequent reflections may have *any* phase angles, and complex algebra is necessary for the summation.

To check your understanding of this, energize the mismatched line of example 2 with a Thevenin source having an open-circuit voltage of 15V and an output resistance of $50\,\Omega$. You should find that the final voltage at the line input and power dissipated in the load are the same, 5V and 1 Watt respectively, but now 2 Watt is dissipated in the source. Only one forward wave, of amplitude 7.5V should exist, and one reflected wave, of amplitude 2.5V. Thus, even though the line is *matched* at the source, more power is wasted to generate the same power in the load!

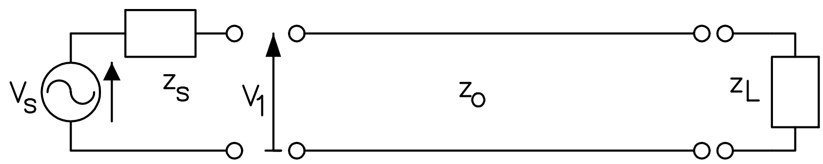
Appendices

A Standing Waves: General Case

Figure A shows a linear Thevenin source generating open-circuit voltage v_s having output impedance z_s connected to the input of a line having characteristic impedance z_o, terminated in a load impedance z_L. All these impedances may be complex, and all may be different.

The line is mismatched at both ends, so reflections will occur at load and source. Consider the source to be switched on at $t = 0$. Then, since no energy has yet been injected into the line, its input impedance will be z_o. By the voltage division theorem, the voltage at the input of the line will be v_1, where

$$v_1 = v_s \frac{z_o}{z_o + z_s} \tag{61}$$



Appendix figure A. A completely general situation, where both source and load are mismatched to a general (lossy) line.

This generates a forward wave, having amplitude v_1, which propagates to the end of the line. When it reaches the load, let its amplitude be v_{f1}. A reflected wave is generated, of amplitude v_{r1} where

$$v_{f1} = v_1 e^{-\gamma\ell} \tag{62}$$
$$v_{r1} = \rho_2 v_1 e^{-\gamma\ell} \tag{63}$$

This first reflected wave travels back down the line to the source, where the mismatch causes it to be reflected with reflection coefficient ρ_1.[14] This reflection generates a second forward wave. The reflected voltage adds algebraically to the initial input voltage v_1 at the source. The current drawn from the source must now also change to accommodate this changed voltage. In effect, the source has "become aware" of the load from the information carried by the reflected wave, and the input impedance of the line must also change.

The second forward wave propagates to the load, where its amplitude is v_{f2}. A second reflected wave is generated, of amplitude v_{r2}, where

$$v_{f2} = \rho_1 \rho_2 v_1 e^{-3\gamma\ell} \tag{64}$$
$$v_{r2} = \rho_1 \rho_2^2 v_1 e^{-3\gamma\ell} \tag{65}$$

[14] An explanation of what happens upon wave reflection from a Thevenin source is given in Appendix C.

This process continues, generating a sequence of multiple forward and reflected waves, which continually diminish in amplitude, since the magnitudes of both ρ_1 and ρ_2 are less than 1. Summing all the forward waves, we get the *total* forward wave amplitude at the load, v_f. Similarly for the total reflected wave, v_r.

$$\text{so } v_f = v_{f1} + v_{f2} + v_{f3} + \ldots \tag{66}$$
$$= v_1 e^{-\gamma\ell} + \rho_1\rho_2 v_1 e^{-3\gamma\ell} + v_1\rho_1^2\rho_2^2 e^{-5\gamma\ell} + \ldots \tag{67}$$
$$= v_1 e^{-\gamma\ell}\left(1 + \rho_1\rho_2 e^{-2\gamma\ell} + \rho_1^2\rho_2^2 e^{-4\gamma\ell} + \ldots\right) \tag{68}$$
$$v_f = v_s \frac{z_o}{z_o + z_s} e^{-\gamma\ell}\left(1 + \rho_1\rho_2 e^{-2\gamma\ell} + \rho_1^2\rho_2^2 e^{-4\gamma\ell} + \ldots\right) \tag{69}$$
$$\text{similarly, } v_r = \rho_2 v_1 e^{-\gamma\ell}\left(1 + \rho_1\rho_2 e^{-2\gamma\ell} + \rho_1^2\rho_2^2 e^{-4\gamma\ell} + \ldots\right) \tag{70}$$
$$v_r = \rho_2 v_s \frac{z_o}{z_o + z_s} e^{-\gamma\ell}\left(1 + \rho_1\rho_2 e^{-2\gamma\ell} + \rho_1^2\rho_2^2 e^{-4\gamma\ell} + \ldots\right) \tag{71}$$

Equations 69 and 71 represent the steady-state, asymptotic values of the wave amplitudes existing on the line when a sufficient time has elapsed for subsequent reflections to be vanishingly small. Note that they are now referenced to the value of v_s, the open circuit *source* voltage. The ratio of these is

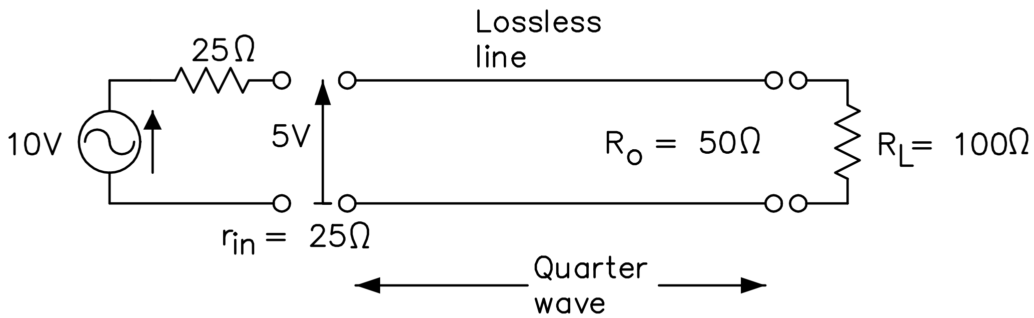
$$\frac{v_f}{v_r} = \rho_2 \tag{72}$$

since all other terms are the same, and cancel. This shows that even in the general case, where *both* source and load are mismatched, *only* the reflection coefficient at the load is required to specify the *ratios* of the steady-state forward and reflected waves, and thus the SWR. This also allows these wave's final, steady-state amplitudes to be calculated directly.

B Example to Illustrate The Derivation

Figure B shows the mismatched quarter-wave line used in example 2 in this paper, but now fed from a linear Thevenin source having an output resistance of $25\,\Omega$. To facilitate comparison with earlier results, the source voltage, v_s, has been set at 10V. Since we energized the line in this first example with 5V, and the input resistance of the line was found to be $25\,\Omega$, we set the open-circuit source voltage at 10V so that *final* conditions on the line will be the same. That is, using the symbols of appendix A, the final, steady-state input voltage on the line will be

$$v_1 = v_s \frac{z_o}{z_o + z_s} = 10\left(\frac{25}{25 + 25}\right) = 5\text{V} \tag{73}$$



Appendix figure B: The lossless, quarter-wave line used in the first illustration fed from a linear Thevenin source.

The final value of the forward wave can be found in terms of the initial conditions by immediately applying equation 69. Rather than doing this, we will follow the progressive buildup of the forward wave to its final value. Again, switch on the source at $t = 0$. The line initially contains no energy, so its input impedance is *not yet* $25\,\Omega$, but $50\,\Omega$, the same as R_o. Hence the initial line input voltage will be

$$v_{f1} = 10\left(\frac{50}{50+25}\right) = 6.6666 \text{ V}. \tag{74}$$

A first forward wave of this amplitude will be generated, and travel to the load. Note that this amplitude is *not* the same as the final, steady-state amplitude, which we determined to be 7.5V.

This wave will be reflected from the load with reflection coefficient $\rho_2 = 1/3$ and travel back to the source. The reflection coefficient at the source, ρ_1, is negative,

$$\rho_1 = \frac{25-50}{25+50} = -\frac{1}{3} \tag{75}$$

But the returning wave, having now travelled forward and back, half a wavelength, has *reversed* its phase, so the amplitude of the second forward wave, and the sum of the two forward waves, will be

$$v_{f2} = \rho_1\rho_2 v_1 e^{-2\gamma\ell} \tag{76}$$

$$\text{where } e^{-2\gamma\ell} = -1 \qquad \text{(2 line lengths travelled)} \tag{77}$$

$$\text{so } v_{f2} = \left(-\frac{1}{3}\right)\left(\frac{1}{3}\right)6.6666\,(-1) = 6.6666\left(\frac{1}{9}\right) = 0.7407\text{V}. \tag{78}$$

$$\text{and the sum } v_{f1} + v_{f2} = 6.6666\left(1+\frac{1}{9}\right) = 7.407\text{V} \tag{79}$$

We see that the forward wave amplitude has increased closer to its final value, but it's not there yet. Adding in all subsequent forward waves, the final forward wave amplitude will be

$$v_f = 6.6666\left(1+\frac{1}{9}+\frac{1}{81}+\ldots\right) = 6.6666\left(\frac{1}{1-\frac{1}{9}}\right) \tag{80}$$

$$v_f = 7.5\text{V} \tag{81}$$

Equation 80 is seen to express the summation represented symbolically in equation 69, and again the result has been obtained using the geometric progression formula. The final forward wave amplitude is thus the *same* as we deduced from simpler arguments in section 4. A similar calculation shows that the final reflected wave amplitude is also correct.

C Reflection from a Thevenin source

What happens in reflection from the source, which contains an impedance *and* a voltage generator, seems to be more perplexing than reflection from an impedance alone.

The sources used in the examples above are linear, Thevenin sources. Such a source has two components, an impedance in series with an "ideal" voltage generator.[15] Thevenin's famous theorem states that

- Any circuit consisting of an arbitrary number of voltage sources and linear impedances connected in an arbitrary fashion, with any two nodes designated as the output terminals, may be represented by an equivalent source having just two components, found as follows:

[15] Thevenin's theorem was historically proved by Gauss, forgotten, and re-discovered by Thevenin and others. Thevenin's original, and most other, proofs are somewhat contrived, and may appear incomprehensible. The simplest and yet completely rigorous proof requires matrix algebra. The only place I know that it appears is in my text, *Theoretical and Computer Analysis of Systems and Networks*, Gary E.J. Bold and Sze M. Tan, any edition, chapter 3, section 6, University of Auckland Physics Department, $20 (NZ)

- The source impedance, z_s, is that impedance measured or calculated between the output terminals when all voltage sources inside are replaced by short circuits.
- The voltage source is that voltage measured between the output terminals of the original network when all external loads are removed.

The voltage source thus obtained (the circle with containing the sinewave with accompanying arrow) has *zero impedance*, which is what is meant by an *ideal source*.

But reflection of any type of wave can *only* occur when the wave encounters an *impedance* discontinuity. Thus reflected waves on the transmission line do not even know that the source voltage generator exists, *because* it has zero impedance.

This may seem strange, because the voltage source is certainly there, so should it not add or subtract somehow with the wave? Yes, but we are working with *linear* circuit elements, and this means that the two processes of wave reflection from the impedance, and subsequent modification of input conditions on the line by the voltage source should be considered *separately* and *then* combined.

The impedance element generates the reflected wave. The voltage source element *adds* its value to that of the reflected wave. This process is used in appendices A and B, and is shown to give the correct results.

D The Direction of Power Travel.

The "forward" and "reflected" voltage waves definitely travel backwards and forwards on the line, and so it seems that power must be also. But we have seen that the "reflected" power is just reversed and sent back towards the load, so that the indicated "forward" power is actually the "forward *plus* reflected power.

We now show that at any time, at any point in the line, the direction of power transfer is always *towards* the load, and its value is constant anywhere on the line.[16]

The voltage and current *on* a coaxial cable give rise to an associated electromagnetic wave *inside* it. Such a wave's power transfer direction *is* unambiguously defined in terms of its constituent electric and magnetic field directions.[17] Figure 3 (a) shows a cut-away section of a coaxial cable, showing both currents and fields. The current is at this moment, and at this point, flowing *into* the page in the centre conductor towards the load, and *out* of the page in the outer shield. We will designate this as the *positive* current direction. The electric field, vector $\mathbf{E}$, can be deduced from Gauss' Law.[18] Here, it is shown pointing radially *from* the centre conductor *to* the outer conductor, implying that the centre conductor is positively charged. The electric field's magnitude is given, in terms of the charge, by

$$E = \frac{\sigma}{2\pi\varepsilon r} \quad (82)$$

$$\text{where } \sigma = \text{the charge per unit length on the centre conductor} \quad (83)$$

$$\varepsilon = \text{the permittivity of the dielectric,} \quad (84)$$

$$r = \text{radial distance from the centre.} \quad (85)$$

This electric field vector is related to the instantaneous voltage between the conductors by[19]

$$V = \frac{\sigma}{2\pi\varepsilon}\log_e\left(\frac{b}{a}\right) \quad (86)$$

$$\text{where } a = \text{the radius of the inner conductor} \quad (87)$$

$$b = \text{the radius of the outer conductor.} \quad (88)$$

[16] I have not seen this proved anywhere else. If *you* have, let me know? *Somebody* has surely done it before!

[17] The equivalence of the "EM wave" and "voltage/current" viewpoints is treated in some advanced texts, see *Communication Circuits*, Lauwrence Ware and Henry Reed, Wiley, third edition, chapter 15.

[18] Gauss' Law is explained in any electromagnetic text, including "*Electromagnetism*", Bold, *Op. cit.*, chapter 4.

[19] The voltage is found by integrating the electric field from the inner to the outer conductor. This is where the logarithm to the exponential base e came from.

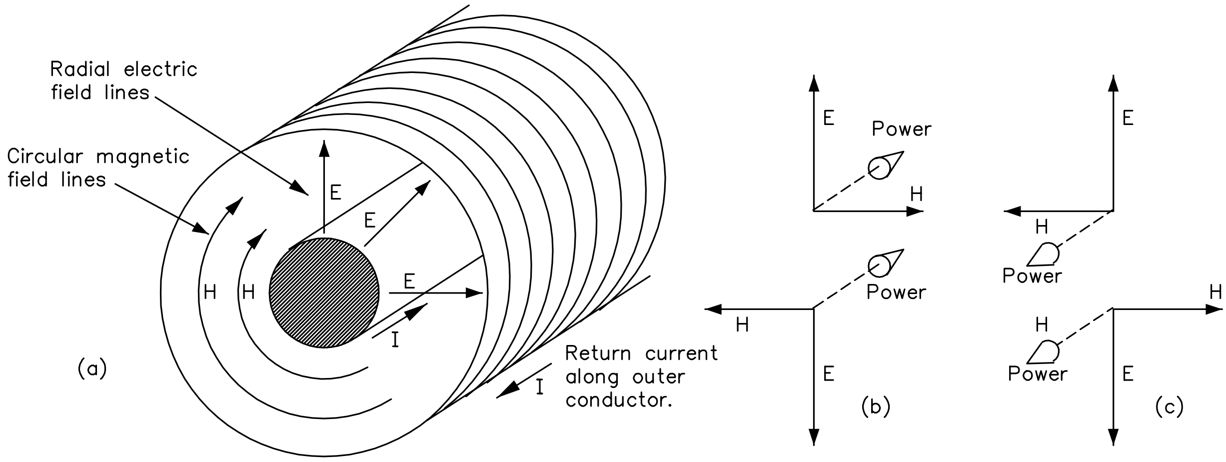


Figure 3: (a) Field lines in a coaxial cable. (b) Field lines in a twin-wire cable.

We see that the magnitudes of the electric field and line voltage are proportional to σ, so V is proportional to E. We will also designate the voltage giving rise to this field as the *positive* voltage direction.

The magnetic intensity field, $\mathbf{H}$, is caused by the current in the centre conductor. Lines of equal magnetic intensity form concentric rings around the centre conductor, inside the cable as shown. Their directions are given by the right-hand screw rule[20] and their magnitude can be calculated using Ampere's circuital Law as

$$H = \frac{I}{2\pi r} \quad (89)$$
$$\text{where } I = \text{current in the central conductor} \quad (90)$$
$$r = \text{radial distance from the centre.} \quad (91)$$

Two such alternating $\mathbf{E}$ and $\mathbf{B}$ fields are *always* associated with an electric field. The magnitude and direction of power travel are given by the Poynting Vector[21]

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} \quad (92)$$
$$\text{where } \mathbf{P} = \text{the vector instantaneous power in the wave} \quad (93)$$
$$\times = \text{the vector cross product.} \quad (94)$$

The magnitude of $\mathbf{P}$ is given by the product of the magnitudes of the $\mathbf{E}$ and $\mathbf{H}$ fields. The direction is found by the right-hand screw rule, and is that in which a right-hand (normal) screw would travel if rotated from the *first* vector towards the *second*.

Figure 3 (b), top, shows the field vectors of figure 3 (a) directly above the conductor. $\mathbf{E}$ is pointing *upwards*, $\mathbf{H}$ points to the *right*. The $\mathbf{E}$ to $\mathbf{H}$ rotation would cause a right-hand screw to travel *into* the page, so this is the direction of power transfer for this current/voltage combination, shown by the dashed line with arrow. However, these vector fields are both alternating sine waves, so they periodically reverse their directions. Half a cycle later the $\mathbf{E}$ and $\mathbf{H}$ arrows will *both* be negative, as shown by figure 3 (b), bottom. The cross-product rotation shows the power transfer vector to be *still* in the *same* direction. We will designate this combination of voltage and current as that of the *forward* wave.

If the current was travelling in the *opposite* direction in figure 3 (a), out of the page, the direction of the $\mathbf{H}$ vector would *reverse*. This situation is shown in figure 3 (c), top. Figure 3 (c), bottom, shows the situation half a cycle later. Now the power transfer vector has *also* reversed its direction, and comes *out* of the page in both diagrams. This combination of voltage and current gives the *reflected* wave.

[20]the rotational direction a right-handed screw must be turned to progress in the direction of the current.

[21]John Henry Poynting published the first, unwieldy derivation of this vector in 1884, but Oliver Heaviside published a simpler one, that normally given now in texts, in 1885. Heaviside's nevertheless requires one inscrutable vector identity. See *Electromagnetism*, Bold, *op. cit.*, chapter 9.

The total voltage at any point in the line, v_t, is the sum of the voltages in the forward and reflected waves, whose magnitudes are related by the voltage reflection coefficient at the load, ρ_2. The maximum and minimum values of the total voltage will occur when v_f is respectively in phase, and in anti-phase with v_f.

$$(v_t)_{max} = v_f\,(1+\rho_2) \quad (95)$$
$$(v_t)_{min} = v_f\,(1-\rho_2) \quad (96)$$

In example 2, section 4, these conditions will occur at the load and the source, respectively. Since ρ is always less than 1, these *total* voltages will be in phase with the *forward* voltage.

The same argument holds for the currents in the forward and reflected waves, since the line current is related to the line voltage by Ohm's Law, via z_o. For simplicity, we assume that z_o is real, as for a lossless line, and we have normalized the lengths of the current and voltage vectors to be the same.[22] The *total* current at these two points will therefore *also* be in phase with the total *voltage.* Figure 4 (a) shows these voltage and current phasor relationships at the source.

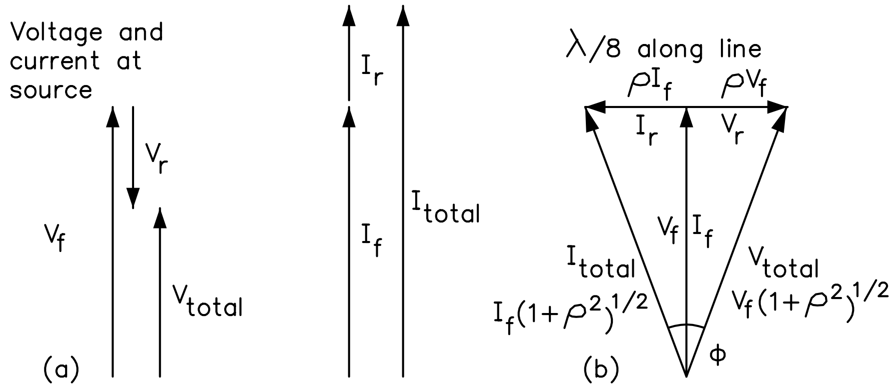


Figure 4: (a) Voltage and current at the source, where the total voltage is *minimum*. (b) Voltage and current one eighth wavelength towards the load.

Hence the $\mathbf{E}$ arrows and $\mathbf{H}$ circles at the source and load will point in the directions shown in figure 3 (a), and the Poynting vector shows that direction of power travel is unambiguously *towards the load* at these points. *At no time* is power flowing in the other direction. In particular, *no* power flows *from the line into the source.*

At other points on the line, the situation is more complex, but we illustrate what happens at a point where the power calculation is geometrically straightforward. Figure 4 (b) shows the situation an eighth of a wavelength towards the load, where the diagrams have been rotated to keep the reference phase of the incident waves vertical, and voltage and current phasors superimposed. The reflected voltage phasor has now rotated 90^o clockwise, or in the retarded direction, while the reflected current phasor has now rotated the same amount the other way.

The *total* voltage and current are no longer in phase. We can calculate the power, P, developed in the line from a geometrical argument. It is well known that power is also given by the product of the *magnitudes* of the voltage and current, and the *power factor*.

$$\text{That is, } P = |V_{total}|.|I_{total}|\cos\phi \quad (97)$$
$$\text{where } \phi = \text{the angle between the total current and voltage phasors.} \quad (98)$$
$$\text{and } \cos\phi = \text{the power factor.} \quad (99)$$
$$\text{so } P = \left(|v_f|(1+\rho^2)^{1/2}\right)\left(|i_f|(1+\rho^2)^{1/2}\right)\cos\phi \quad (100)$$
$$= \left(|v_f|(1+\rho^2)^{1/2}\right)\left(\frac{|v_f|}{R_o}(1+\rho^2)^{1/2}\right)\cos\phi \quad (101)$$
$$= \frac{|v_f|^2}{R_o}\left(1+\rho^2\right)\cos\phi \quad (102)$$

[22] although these results are true for *any* line. The algebra just becomes more tedious.

$$\sin(\phi/2) = \frac{\rho}{(1+\rho^2)^{1/2}} \tag{103}$$

$$\text{and } \cos\phi = 1 - 2\sin^2(\phi/2) = 1 - \frac{2\rho^2}{1+\rho^2} \tag{104}$$

$$\cos(\phi) = \frac{1-\rho^2}{1+\rho^2} \tag{105}$$

$$\text{so } P = \frac{|v_f|^2}{R_o}\left(1+\rho^2\right)\frac{1-\rho^2}{1+\rho^2} \tag{106}$$

$$P = \frac{|v_f|^2}{R_o}\left(1-\rho^2\right) \tag{107}$$

$$\text{Checking, for example 2, } P = \frac{7.5^2}{50}\left(1-\frac{1}{3^2}\right) = 1 \text{ Watt,} \tag{108}$$

which is the same answer as obtained earlier by power conservation. In fact, this relationship holds all along the line, but is tedious to show in the general case. Since ρ_2 is always less than one, the total voltage on the line will *always* have its largest component in the direction of v_f.[23] It follows that the electric field will also have its largest component in this direction. Similarly for the magnetic field.

It follows that the *average* power vector $\overline{\mathbf{P}}$, evaluated over a complete cycle, will *also* always point towards the load, wherever, and whenever it is evaluated.

We now confirm that the *total* power carried by the electromagnetic wave inside the line is the *same* as that computed from the voltage and current phasors. The electromagnetic power inside the line will be the value of vector $\mathbf{P}$ integrated over the area of the line between the inner and outer conductors. Because both The E and H wave magnitudes vary with radius, an integration is necessary. Let V and I be the *instantaneous* magnitudes of the total voltage and current on the line. Consider the contribution to the power, dP, of the field in an annulus around the line centre, at radius r, and infinitesimal width dr.

$$dP = EH.2\pi r.dr \tag{109}$$

$$\text{substituting for } E \text{ and } H,\ dP = \left(\frac{\sigma}{2\pi\varepsilon r}\right)\left(\frac{I}{2\pi r}\right) r.dr \tag{110}$$

$$\text{Integrating, the total power } P = \frac{\sigma I}{2\pi\varepsilon}\int_a^b \frac{dr}{r} \tag{111}$$

$$P = \frac{\sigma I}{2\pi\varepsilon}\log_e\left(\frac{b}{a}\right) \tag{112}$$

$$\text{but from equation 86, } V = \frac{\sigma}{2\pi\varepsilon}\log_e\left(\frac{b}{a}\right) \tag{113}$$

$$\text{where } V = \text{the total voltage across the line at this point.} \tag{114}$$

$$\text{therefore } P = VI \tag{115}$$

Thus the total instantaneous power carried by the electromagnetic wave *inside* the line is the *same* (as expected) as the power given by the product of voltage across, and the current through, the line. This relationship holds at any point, at any time, on the line, though V and I will vary along the line, and are related by equation 45, through $V = I\,z_{in}$.[24]

To summarize, this appendix shows that

- The *direction* of power travel on the line is *always* from the source to the load. *No* power travels in the other direction, *even though* the reflected wave *does* travel from the load to the source.
- The *magnitude* of this power is invariant along the line, and can be calculated *either* from the voltage and current on the line, or from the power carried by the electromagnetic wave *inside* the line.

[23]strictly speaking, we should use the *magnitude* of ρ_2, but in the examples worked ρ_2 is always real, so we will stick with this. The result in the general case can be shown to be the same.

[24]although z_{in} has been quoted as the *input* impedance at the end of the line, in fact this same relationship gives the impedance at *any* point *in* the line. See any text on transmission line theory for verification.

Conclusions

- A Bruene power/SWR meter's operation can be explained from two points of view. If it is connected directly to, or in a transmission line, its readings can be interpreted as powers in "forward plus reflected" and "reflected" waves. Alternatively, it can be considered as measuring the departure of the impedance seen looking out from its output terminals from a pure resistance of $50\,\Omega$. This holds whether it's connected to, or in, a line, or directly to a lumped impedance.
- On a transmission line, unless the source can be replaced by a *linear* source impedance matched to the line characteristic impedance, reflections *always* occur at the source.[25]
- But rf output stages operating in class B or C are non-linear, and can't be replaced by Thevenin sources for input-end reflection coefficient calculation.
- Regardless of whether the source is matched or not, the SWR is determined *only* by the reflection coefficient at the load.
- The *first* forward wave has the *same* amplitude as the initial line input voltage. But this is *not* the same as the *final, total* forward wave amplitude. Similarly for the reflected wave.
- The final, steady-state amplitudes of the forward and reflected waves result from the summation of (in principle) an infinite number of reflected waves.
- Steady-state conditions on the line require time to build up. That is, energy has to travel to and fro on the line a sufficient number of times for reflections to become vanishingly small.
- *All* line conditions, including the way in which the waves build up, can be deduced even for a lossy line, a mismatched load and a mismatched *linear* source, *provided* that the *reflection coefficient of the source* is correctly included, and we use the Thevenin equivalent representation of the source.
- Electromagnetic theory shows that *all* power travels from the source to the load, anywhere on the line, at all times. This power is the same at *all* points on the line. No power flows in the other direction.

[25] in general, if the line is lossy, its characteristic impedance will be complex, so a (complex) conjugate match is required. Of course, this is unlikely to be the case for any real transmitter!