SWR

—What it means in practice

When amateurs get together, and the talk turns to antennas, it is not long before the magic phrase, "SWR", is heard. But just what is SWR and how important is it in practice? This article looks at the subject in practical terms, at a level which everybody should understand.

by PHILIP WATSON VK2ZPW

It is an unfortunate fact that SWR, along with antenna and transmission line theory in general, is one of the most misunderstood subjects in the whole of amateur radio. It boasts as many myths and old wives tales as does pregnancy and childbirth, some of them perpetuated by supposedly reputable text books.

In trying to get a mental picture of what is, admittedly, an extremely complex subject it is often a help to start with a theoretically perfect situation, against which we can compare the usually imperfect practical situation.

The antenna

Let's start with the antenna. In any transmitter installation, this has to satisfy two basic requirements. One is to radiate the RF energy fed to it by the transmitter in the most efficient manner possible. The other is to present the transmitter with the correct load in order that the transmitter may deliver the level of RF power which the designer intended.

While both are important, the second requirement is, in many ways, the more important one. Suppose we have a typical commercial transmitter designed to deliver 10W into a 50Ω load. If we connect a pure (ie, non-reactive) 50Ω resistor directly across the antenna terminals or socket at the set, and energise the transmitter, it will deliver 10W to the resistor, which will appear as heat.

If we were to substitute some other value of resistor a number of things could happen. The most likely one is that the transmitter would no longer deliver its 10W. By how much it would fall short would depend on the error in the load value and the design and tolerance of the particular transmitter output stage.

Another possibility, again depending on the output stage design, is that it would try to deliver more than its rated power, but run into overload in the process, and destroy itself and other sections as a result. Fortunately, most commercially designed transmitters are well protected in this regard, but there is no point in taking unnecessary risks.

But, this risk aside, we should make every endeavour to present the correct load to the transmitter simply to ensure that it delivers the maximum power for which it was designed. On the other hand, there is little to be gained by simply feeding this energy into a resistor; it will radiate very little of the RF energy and waste virtually all of it as heat.

So we replace the resistor with an antenna and, fairly obviously, this antenna should look (to the transmitter) like a 50Ω resistor if it is to deliver maximum power. Assuming the antenna is resonant at the transmitter frequency, and fed at the right point, it will look like a resistor. If it is not resonant it will exhibit either a capacitive or inductive component, according to which way it is off resonance.

Assuming that it is resonant, the next question concerns the value of resistance it presents to the transmitter. And this is where the going gets tough because, in other than a very few clearly defined cases, this is very largely an unknown or, at best, "guesstimated" value.

We can, for example, nominate the resistance at the centre of a half wave dipole as being in the region of 72Ω, while a folded version of the half-wave dipole will have four times this impedance, or 288Ω (often mentally rounded off to 300Ω). A number of factors can cause minor variations to these values, such as the diameter of the elements, relative to their length, space between folded dipole elements, etc.

A more controversial value is that for the popular quarter-wave ground plane. For years it has been stated, in many popular amateur textbooks, that this is approximately half the value of the simple dipole, or 36Ω (in fact, various values have been quoted between 30 and 36Ω). This figure appears to have been based on a theoretically calculated value for a quarter-wave radiator working against an infinitely large, perfectly conducting ground plane.

Some of these text books even went so far as to describe matching devices ("Q" sections, etc) which would match this value to the popular 50Ω coax cable and transmitter load requirements. As anyone who has tried to make one of these matching systems work, or who has attempted to confirm this figure with an impedance bridge will testify, the reallife ground plane, using four quarter-wave radials, is a vastly different device.

Strangely enough, the true situation has been known for many years. At least as early as 1962, and possibly earlier, the R.S.G.B. Handbook (page 365) stated: "The radiation resistance of a ground quarter wave aerial is 35Ω, but that of a ground-plane is less than 20Ω." More recently other authorities have been emphasising this same point but, unfortunately, old ideas die hard, and the point needs to be emphasised a good deal more strongly if the error is to be corrected.

One of the most recent articles on the subject, and one of the most detailed, is that by Guy Fletcher, VK2BBF, which appeared in the August 1984 issue of "Amateur Radio", the official journal of the Wireless Institute of Australia. This, and two articles which follow in subsequent issues, are recommended to anyone wishing to get a more detailed picture of the antenna scene generally.

But this is a separate argument. The point we set out to make is that, apart from these few simple designs, it is exceedingly difficult to nominate the feed point impedance of an antenna. We do know the general effect of many design factors; that, for example, the addition of director or reflector elements to a dipole will lower the impedance. But by how much is another matter.

So we are faced with the situation that the impedance of all but the simplest antenna systems is largely an unknown quantity. With experience we can make a rough estimation that it will be between this and that figure, or below some other figure, but beyond that we must resort to some form of measurement or "suck-it-and-see" approach.

SWR meters

One such approach involves the use of a standing wave ratio meter, or SWR meter. But it would be premature to go into details at this stage. We need to talk about SWR in some detail first.

So far we have considered only those situations where a resistor or an antenna — which looked like the same resistor — was connected directly across the transmitter output terminals. Apart from a few special cases, feeding an antenna in this manner is not very practical. We need to locate the antenna as high as possible and clear of objects which might shield it, while we need to put the transmitter in a convenient indoor working location, some distance away.

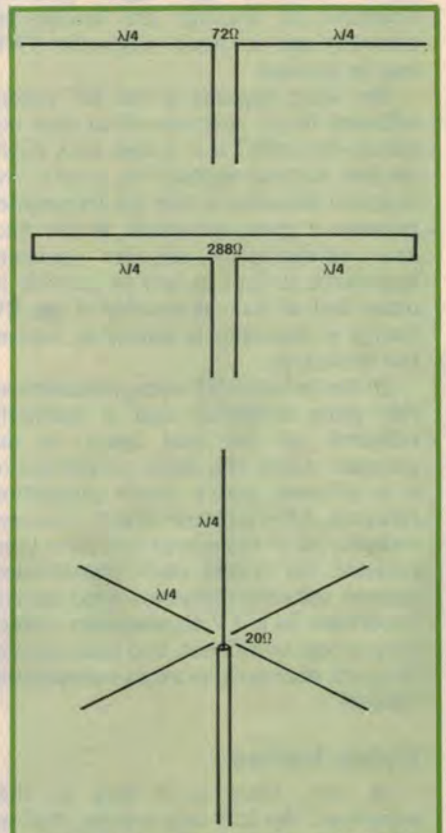
And, to couple the two together, we need a special kind of cable; one that will convey the transmitter output to the antenna with minimum loss and which, in itself, will not radiate any significant amount of this energy into a shielded environment, where much of it would be wasted.

There are two types of cable commonly used by amateurs, the open wire line and the coaxial cable. The open wire line can be homemade, has very low losses, and can be made to have any impedance characteristic over a wide range. On the other hand it can be awkward to install and is much less popular than it once was.

Coaxial cable is a commercial product, with somewhat higher losses, and is commonly available in two popular impedance values: 50 and 72Ω. It is reasonably flexible and relatively easy to install. For most of our discussion we will assume the use of coaxial cable, although most of the points would be just as valid for open wire lines.

Characteristic impedance

Undoubtedly the most important single characteristic of a coaxial cable, for the beginner to understand, is its characteristic impedance, typically 50 or 72Ω, as already mentioned. This is not an easy concept to grasp and the beginner may have to content himself with accepting some basic statements at their face value, at least initially.



Three basic antennas for which the feedpoint impedance is known with reasonable accuracy.

Coaxial cable consists of two conductors, one within the other, and insulated from each other. A common form uses solid or stranded wire as the central conductor, copper braid as the outer conductor, and a polythene insulating material between them.

The characteristic impedance is a factor of the inductance of the two conductors, relative to the capacitance between them, per unit length. These factors, in turn, are determined by the physical characteristics of the components; the inductance by the cross sectional area of the conductors, and the capacitance by their area relative to each other, the distance between them, and the dielectric constant of the insulating material. The length of the line is not a factor.

The effect of this inductance/capacitance relationship is to establish an equally firm relationship between the voltage and current of RF energy travelling up the line. This relationship is exactly the same as would have occurred across and through a pure resistor having the same value (say 50Ω) as the characteristic impedance of the cable.

It may help to grasp this concept if one is to visualise a very short burst of RF energy transmitted up the line; so short that its trailing edge has left the transmitter long before its leading edge has reached the load at the far end. Thus, something in the manner of a fired projectile, or even a thrown tennis ball, it is in a kind of limbo; while influenced initially by the manner of its launch its subsequent movement is largely a factor of its environment. And it knows nothing about what lies in store for it at the end of its journey.

We can carry the analogy a little further. If the tennis ball ultimately hits a brick wall it will bounce off (or be reflected) simply because the brick wall represents a gross mismatch to the manner in which energy is stored in the moving ball. A softer object, such as a bale of hay, may well have absorbed all the energy with no bounce (or reflection).

The same applies to our burst of RF energy. When it reaches the end of the cable it will need to meet exactly the right load if all its energy is to be dissipated in that load. And it doesn't take much imagination to conclude that the load should look like (in this case) a 50Ω resistor.

If it encounters any other value then only part of the energy will be absorbed by the load, and the remainder will be reflected down the line in the direction of the transmitter. And this is what creates what are called "standing waves" on the transmission line.

Standing waves

In greater detail, the standing waves are actually peaks of voltage between the conductors, or peaks of current through the conductors, which occur at regular intervals along the line. They occur at those points where (say) the voltage of the outgoing energy encounters voltage of reflected energy which is exactly in phase with it. Similarly for the current peaks.

The position of each peak is fixed and will always be one half wavelength away from its neighbour. Exactly between each peak, ie, one quarter wavelength away, will be a dip or voltage minimum, and it is the ratio between these two voltages which constitutes our "standing wave ratio" or SWR. (Note: wavelengths in coaxial cable will be physically shorter than in free space, according to the characteristics of the insulating material. A factor of 0.66 is typical, commonly referred to as the "velocity factor".)

In the theoretically perfect situation, where the cable is correctly terminated, all the energy is absorbed by the load, there will be no reflected wave, and the voltage and current values will remain essentially constant along the length of the line. Such a situation is said to constitute a "flat" line.

By now the reason for our interest in the SWR should be apparent. Because it occurs only if there is a mismatch, and its value is directly related to the degree of mismatch, its measurement provides a very useful "suck-it-and-see" approach to ensuring that the transmitter is presented with its correct load.

In greater detail, an SWR of (say) 2 to 1 will mean that the load is in error, relative to the cable impedance, by this ratio. But it cannot indicate in which direction the error lies. Assuming a 50Ω cable the 2 to 1 error could mean that the load is half (25Ω) or twice (100Ω) the correct value. Note, however, that this relationship is true only when the load is purely resistive.

Considering all the foregoing, and with the benefit of hindsight, one wonders whether the term "SWR", to some extent, might be misleading; and that some other term, like "mismatch ratio", might not have been a better choice.

But it is essential to keep one very important point in mind at this stage of the discussion. The existence of standing waves, in itself, is only a secondary problem. It is a useful measurement only because it tells us whether the transmitter is being correctly loaded or not and that our efforts should be directed to correcting this aspect of the problem. Whether we correct the SWR in the process may not even matter. Let us consider a practical example.

Suppose an SWR reading indicates quite clearly that there is a serious mismatch between antenna and cable. We have two options: either fit some kind of matching device between the antenna and the cable so that the antenna now looks like the correct value, or fit a matching device between the transmitter and the cable so that the transmitter sees a correct load.

In theory the first option is the preferred one, since we not only present the transmitter with its optimum load, but we eliminate the standing waves at the same time. In practice, however, the second option may well be very much more practical and convenient. It will have achieved the same primary objective of loading the transmitter correctly and in many cases the SWR can be ignored.

But what happens to the RF energy reflected by an antenna which does not match the cable? If it is sent back down the line, is it not wasted? No, it isn't. The practical situation is that the transmitter presents a gross mismatch to the line, and deliberately so. Its (source) impedance is kept as low as possible in order that as little as possible of the RF energy it generates is wasted as heat in the final stage.

So the reflected RF energy encounters this gross mismatch and is promptly reflected up the line again to the antenna, where the major proportion of it is radiated and a minor proportion reflected. After a couple of such journeys virtually all of the energy will have been radiated. (In typical audio transmission systems the time delays involved are not important. In a TV transmission system they can be significant, and more careful design is necessary to avoid transmitting "ghosts".)

Cable losses

In fact, there is a flaw in that argument. We can only assume that no energy is wasted if we ignore the inherent losses in the cable. All cables have some losses, and these increase with frequency. For example, a popular foam filled coax, RG8U, has a loss of 0.9dB/30m at 30MHz which rises to 3.5dB/30m at 400MHz. The presence of such losses means that any signal which has to traverse the line more than once will suffer additional losses on each excursion.

So we have to concede that, in practice, standing waves do create some loss. But how much, and how important is it? If we assume a 3dB loss in a cable system which is correctly terminated, ie, no standing waves, then an SWR of 3 to 1 will add a further 1dB loss. The accompanying graph indicates the additional loss for a wide range of basic cable losses and SWR values.

At 450MHz, using RG8U cable, with a run approaching 30m, a loss of 3dB could be expected and, if it had to be tolerated, then anything which would minimise further losses would be worth considering. This is a case where, all else being equal, correction at the antenna might be preferable to that at the transmitter.

At lower frequencies losses become less important. At 150MHz, RG8U wastes only 2dB/30m (an additional 0.8dB for a 3:1 SWR), and at 30MHz 0.9dB (plus 0.48dB for 3:1 SWR).

So, hopefully, that should put the SWR problem into some kind of perspective. But there are other misconceptions which we might perhaps comment upon. One is that the reflected energy finds its way back into the final stage and overheats it. Wrong!

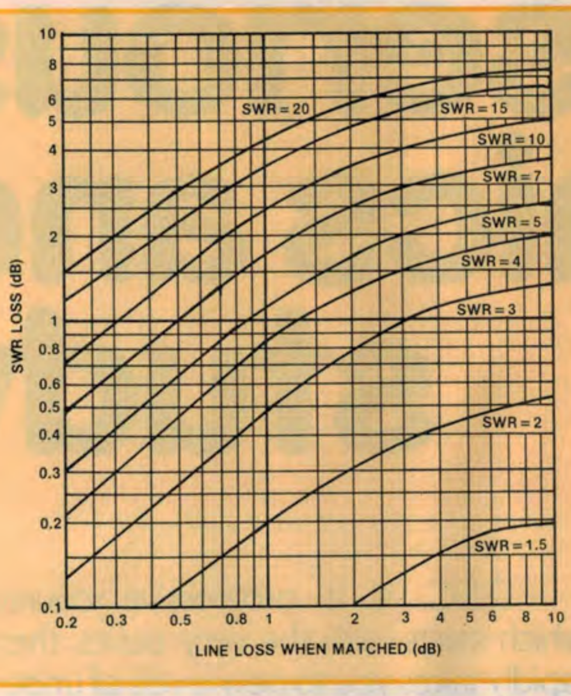
It is true that a transmitter working into a transmission line with a high SWR may show signs of distress. But the distress is not due to the SWR; rather it is due to the incorrect load at the antenna into which the transmitter is trying to work.

Cable length?

Another popular furphy claims that the length of the cable is critical; that it must be an exact multiple of a half wavelength long if the transmitter is to be properly loaded — and the standing waves eliminated — even when the antenna is presenting a proper load.

The truth is that, if the load is correct, then this value will be "seen" at the other end of the cable regardless of its length. If the load is incorrect, then this value will be seen at half wavelength intervals along the cable. But since it is wrong anyway there seems little point in trying to reproduce it.

In fact, in such circumstances, the length of the line can be critical for a



This graph clearly indicates the amount of additional loss, due to SWR, which occurs for various values of pure cable loss ie, loss when correctly terminated.

quite different reason. Between the half wavelength points the cable will exhibit a range of impedances, one of which may match the transmitter. So, by adjusting its length the cable may be made to act as a matching transformer, and load the transmitter correctly.

But don't try to do it using the SWR meter because altering the line length will have no effect on the SWR. If this trick is to be employed other measurements must be used, such as that from a field strength meter at a fixed distance from the antenna.

So, after all that, what is the role of the SWR meter? Well, it obviously isn't the universal answer to all antenna/transmission line problems. On the other hand, if it is the only instrument available it can be quite useful. An important point to realise is that, while it can indicate that there is something wrong with a particular set-up, it cannot indicate what is wrong.

Thus an antenna may present the wrong load for a number of reasons. It may not be resonant, the design may be wrong or may have been misinterpreted by the constructor, or the matching device, if one is used, may be incorrect. Alternatively, the cable impedance may be other than that claimed. (There is the story, well authenticated, about the Sydney disposals dealer who could supply either 50Ω or 75Ω cable at a very attractive price, both off the same reel!)

In other words, when the SWR meter indicates that there is something wrong, the important thing is to make a systematic approach to finding out what it is. For example, terminating the cable in a good dummy load having the same resistance will quickly indicate whether or not the cable is at fault. If it is not, the antenna is the next obvious suspect.

Exactly what needs to be done, or can be done, to change the antenna's impedance will, of course, depend on the particular type of antenna and what is physically convenient or practical. But, whatever the approach, the SWR meter can be used to monitor the effect of the changes or adjustments.

Finally, one more controversial point. Just where should the SWR meter be connected in the line: at the transmitter end or the antenna end? Some authorities are adamant that it should be at the antenna end, while others are equally emphatic that this precaution isn't necessary.

While, in theory, it can be shown that the antenna is the right place to make this measurement, the practical situation is that this is seldom a very convenient, or even feasible, arrangement. So, in practice, most people tend to make it at the transmitter end. (Where an antenna tuning unit, or other matching device is used at the transmitter, fitting the meter between the two is a perfectly legitimate way of determining when the tuning unit is presenting the correct load to the transmitter.)

The main objection to measurements made at the transmitter end is that the cable losses will mask the true ratio, the forward signal having been attenuated before it was reflected, and the reflected signal attenuated again on the way back. Depending on the severity of the losses, the user may obtain a reading below what he has set as an acceptable maximum when, in fact, the true value is appreciably higher.

Unfortunately, cable losses become worse as the frequency increases and, in the 420-450MHz (70cm) band this problem could be very real. So, be prepared to work at the antenna unless the coax line can be kept short. In cases like this it is sometimes better to move the transmitter close to the antenna, and use a much shorter line.

And so to sum up: The most important characteristic of an antenna system is to present the transmitter with its optimum load. An SWR measurement can indicate whether this is happening and, if not, the degree of error. It is valuable primarily for this reason, the standing waves in themselves being relatively unimportant.

So let's keep things in perspective.