8

Metrology for Vector Network Analyzers

Roland G. Clarke and Nick M. Ridler

8.1 Vector Network Analyzers

8.1.1 The role of a network analyzer
8.1.2 Scattering parameters
8.1.3 VNA systems
8.1.4 Extender heads
8.1.5 Good practice tips

8.2 Metallic waveguides

8.2.1 Basic properties
8.2.2 Dispersion
8.2.3 Standardized sizes and frequency ranges
8.2.4 Flanges / Interfaces
8.2.5 Good practice tips

8.3 Calibration Standards and Methods

8.3.1 Calibration: general principles
8.3.2 Types of standard
8.3.3 One-port calibration
8.3.4 Two-port calibration
8.3.5 Good practice tips

8.4 Errors and Uncertainties

8.4.1 Main sources of measurement error
8.4.2 Typical sizes of errors
8.4.3 Connection repeatability
8.4.4 System verification
8.4.5 Good practice tips

8.5 Concluding remarks

References
The Vector Network Analyzer (VNA) has been a feature of radio-frequency (RF) and microwave measurements since the 1960s. Early instruments were often physically large, requiring plenty of manual adjustment to obtain good measurements, and they were generally limited to frequencies well below 20 GHz. However, rapid advances in digital electronics and signal processing, combined with significant progress in microwave signal sources and components, has seen the development of today’s compact, highly automated VNAs. The VNA is now widely used to measure passive and active microwave components, and to characterise high-frequency materials.

During the past twenty years, a range of extension modules have been developed by various manufacturers which effectively increase the measurement capability of VNAs to much higher frequencies [1-3]. 'State of the art’ network analyzers equipped with suitable high-frequency extension modules are now capable of measurements at 1 THz. It is these systems which are likely to be found in a THz measurement laboratory.

In this chapter we will introduce the typical, commercially-available, VNA equipment adapted for millimeter-wave, submillimeter-wave and THz measurements (i.e. used at frequencies ranging from 100 GHz to 1 THz, or thereabouts). Other VNA systems are in use in some laboratories which adopt a different arrangement or measurement technique (see, for example, [4, 5]). These other VNA systems will not be discussed further in this chapter. We shall begin by explaining what the VNA actually measures, and then move on to discuss some of the metrology issues which need to be considered by the laboratory practitioner.

8.1 Vector Network Analyzers

8.1.1 The role of a network analyzer

As the name suggests, the network analyzer is an instrument which provides information about the scattering of signals\(^1\) incident upon a network. In this context, a ‘network’ is any circuit, device or more generally, any physical artefact, to which electromagnetic signals may be applied through one or more ‘ports’. The notion of a port is significant because the underlying basis of network analysis assumes that, regardless of what the network consists of, it is connected to the outside world by means of a well-defined system for guiding the incident and emerging signals. In the case of millimeter-wave, submillimeter-wave and THz frequencies, this guiding structure is most likely to be a metallic waveguide.\(^2\) The network that is to be measured is usually referred to as the ‘device-under-test’ (DUT).

Since the ports are physically realised by a guiding structure, both an impedance (characteristic of the guiding system) and a precise location (i.e. plane) are associated with each network port.

When an electromagnetic wave propagates within a guiding structure, the ratio of electric (voltage) and magnetic (current) portions of the propagating disturbance is a property of the structure itself (geometry, materials, etc). This ratio is considered to be a characteristic impedance (denoted \(Z_0\)) associated with the guiding structure at a given frequency of operation [6]. If, at some point, the propagating wave encounters a different impedance, some (possibly all) of the wave energy is reflected. We may think of this process as a ‘scattering’ of the incident signal, and it may be shown that the amount of reflection (the reflection coefficient) is directly related to the relationship between the two impedances. For the purpose of measurements, the impedance associated with each port is often termed the ‘reference impedance’, and is often (though not necessarily) the same for each port of a given network.

\(^1\) We use the term ‘signal’ here to mean ‘measurement signal’, which is ordinarily an unmodulated, pure sinusoid, rather than an information-carrying, aperiodic signal.

\(^2\) See the following section for a more detailed discussion of metallic waveguides.
For some types of guiding structure (for example, a co-axial cable) it is relatively straightforward to define this impedance, especially when a transverse electromagnetic wave is propagating in a low-loss ‘transmission line’. In the case of metallic waveguides, the characteristic impedance is not so readily defined. However, in practice this does not matter too much, providing it is recognised that any discontinuity in the electric/magnetic ratio encountered by the propagating wave, will generally lead to reflections. It is not necessary to know (or assign) a numerical value to the characteristic impedance in order to make meaningful measurements of the reflection of signals incident on a given network. ³

From a measurement perspective, the physical location of the port – the ‘reference plane’ – is also important. This is because the incident and emerging signals are travelling waves, and their phase relationships will be a function of position. By measuring at different locations along a waveguide, different phase relationships will be observed. For convenience, the reference planes for practical measurements are usually taken to be at the junctions between two waveguide sections.

By applying continuous, sinusoidal, electrical stimulus signals to the network (DUT), the VNA measures the incident and emerging signals at each port. The results are presented as ratios of emerging signals to incident signals. Thus, the role of the network analyzer is to measure the scattering of signals incident on each port of the DUT. The vector network analyzer is so-called because it is capable of measuring both the magnitude and the relative phase of these signals.⁴

It is important to realise that the magnitude and phase relationships of the incident and emerging signals are generally frequency-dependent, and therefore the VNA usually measures the scattering of signals as a function of frequency.

8.1.2 Scattering parameters

The notion of ‘reflection coefficient’ can be extended if we consider the alternative ways in which the energy in a guided electromagnetic signal may be distributed, or ‘scattered’.

First, let us consider a network with only a single port (see figure 8.1 (a)). The energy associated with an incident signal might be completely absorbed by the network, either by thermal dissipation or by radiation.⁵ Certainly, this must be the case if there is no steady-state⁶ impedance discontinuity experienced by the incident wave as it enters the network port.

Alternatively, the energy may be partially – or wholly – reflected back to the original source. By energy conservation, the sum of the reflected and absorbed energies must be equal to the original incident energy. We may also think in terms of time-averaged rates of energy transfer, or average power. In which case, the above statement also holds true, and the sum of the average reflected and absorbed signal powers must be equal to the average of the incident signal power.

This idea can be extended to two-ports (see figure 8.1 (b)), and indeed, n-ports. For a two-port network, there are three possible ‘paths’ for the incident signal power. Once again, the incident signal might be absorbed or reflected by the network, but it might equally well be transmitted through the

³ There are various ways to define ‘impedance’ in a metallic waveguide. These generally differ by a constant term, and since it is ratios of impedance that determine reflections, the choice of definition is not important.

⁴ By contrast, a ‘scalar analyser’ only measures the magnitudes of the incident and emerging signals.

⁵ Internal dissipation is indistinguishable from radiating the energy away, as far as this network’s input port is concerned, and so both may be simply treated as ‘absorption’.

⁶ The use of the term ‘steady-state’ is necessary because in some networks, once the continuous, sinusoidal electromagnetic signals have settled to a final state, an impedance discontinuity which might have been encountered by an initial disturbance may effectively disappear. This is the basis of so-called ‘impedance transformers’ (for example, those which rely on quarter-wavelength lines) [7].
network and emerge from the second port. Thus, a ‘transmission coefficient’ may also be considered, which is related to the ratio of the signal emerging from the second port to the signal incident at the first port. Again, energy conservation requires that the sum of the average reflected, absorbed and transmitted powers be equal to the average incident signal power.

![Diagram of a one-port network](a)

![Diagram of a two-port network](b)

Figure 8.1: (a) A ‘one-port’ network; (b) a ‘two-port’ network. Incident energy ‘scatters’ by reflection, internal absorption or by transmission to another port.

By considering the incident and emerging signals from each port of a network, we can now define a set of scattering parameters (or scattering coefficients). In figure 8.2 below, the incident signals on each port are denoted as ‘\( a_1 \)’ and ‘\( a_2 \)’, whilst the emerging signals are ‘\( b_1 \)’ and ‘\( b_2 \)’.

![Diagram of a general two-port network](general_two_port_network)

Figure 8.2: A general two-port network, with incident signals \( a_n \) and emerging signals \( b_n \)

The ratio of the signals \( b_1 \) and \( a_1 \) gives the scattering coefficient for a signal which is both incident on Port 1, and also emerging from Port 1. This, of course, simply represents the reflection of a signal arriving at Port 1, providing that there are no other incident signals on the network which could potentially contribute to the signal emerging from Port 1. Therefore, this scattering parameter is defined only when \( a_2 \) is zero.

If we denote this parameter, ‘\( S_{11} \)’, then we can write

\[
S_{11} = \frac{b_1}{a_1} \bigg|_{a_2=0}
\]  

(8.1)

The first subscript for the scattering parameter designates the port from which a signal emerges, whilst the second subscript designates the port on which the signal is incident.
By taking similar ratios of other combinations of the ‘a’ and ‘b’ signals, four scattering parameters may be defined for a two-port network. These parameters describe the reflections for each port and the transmission between each port. It is not necessary to have a separate coefficient for absorption, since the information about reflection and transmission alone conveys a complete description of the scattering of incident signals. Such scattering parameters are often called simply ‘S-parameters’ and as such, they are fundamental to VNA measurements. Table 8.1 shows these scattering parameters and their corresponding physical interpretation:

<table>
<thead>
<tr>
<th>Scattering parameter</th>
<th>Relationship to ‘a’ and ‘b’ signals</th>
<th>Physical meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{11}$</td>
<td>$S_{11} = \frac{b_1}{a_1}</td>
<td>_{a_2=0}$</td>
</tr>
<tr>
<td>$S_{12}$</td>
<td>$S_{12} = \frac{b_1}{a_2}</td>
<td>_{a_1=0}$</td>
</tr>
<tr>
<td>$S_{21}$</td>
<td>$S_{21} = \frac{b_2}{a_1}</td>
<td>_{a_2=0}$</td>
</tr>
<tr>
<td>$S_{22}$</td>
<td>$S_{22} = \frac{b_2}{a_2}</td>
<td>_{a_1=0}$</td>
</tr>
</tbody>
</table>

Table 8.1: Physical meaning of two-port scattering parameters.

The individual scattering parameters are defined under the conditions that a signal is incident upon one port at a time. However, providing the network is linear, the emerging signals may consist of a sum of contributions from the incident signals at each port. We can therefore write two linear equations relating the emerging signals, the scattering coefficients and the incident signals:

\[
\begin{align*}
  b_1 &= S_{11}a_1 + S_{12}a_2, \\
  b_2 &= S_{21}a_1 + S_{22}a_2.
\end{align*}
\]  
\[ (8.2) \]  
\[ (8.3) \]

These equations may also be written in matrix form,

\[
\begin{bmatrix}
  b_1 \\
  b_2
\end{bmatrix} =
\begin{bmatrix}
  S_{11} & S_{12} \\
  S_{21} & S_{22}
\end{bmatrix}
\begin{bmatrix}
  a_1 \\
  a_2
\end{bmatrix}
\]  
\[ (8.4) \]

For a network with two or more ports, the scattering parameters are often referred to as the ‘scattering matrix’, or ‘S-matrix’. We must now look more closely at the way in which the ‘a’ and ‘b’ signals are defined.

Firstly, the ‘a’ and ‘b’ signals represent travelling electromagnetic waves with both magnitude and relative phase. Therefore, they must be treated as complex (phasor) quantities, and the scattering parameters will also be complex quantities.

The magnitudes of the ‘a’ and ‘b’ signals are defined in terms of the square-root of power, rather than voltage (E field), current (H field) or simply power. This may seem confusing at first, but there are good reasons for adopting this convention. One of the reasons is that the scattering parameters should provide an immediate insight into the behaviour of a network.

Consider a network which is reciprocal, such that the ratio of power transmitted from Port 1 to Port 2 is identical to the ratio of power transmitted from Port 2 to Port 1. Intuitively, we might expect the
forward and reverse transmission scattering coefficients to be the same. But if the reference impedance is different for the two ports, only those ratios related directly to power would be the same. The ratios of voltages or currents, for example, would be different in each direction.

However, if the propagating voltages and currents are normalised to the square root of the reference impedance, we obtain the desired results. Dividing voltages by the square-root of impedance yields a parameter with the dimensions of the square-root of power; similarly, multiplying a current by the square root of impedance also yields a parameter in terms of the square-root of power. Therefore, the choice of square-root of power (rather than simply power) for the ‘a’ and ‘b’ signals arises from the fact that the scattering coefficients are the same, whether derived from voltages or currents, and also because the scattering matrix is symmetrical for reciprocal networks.

A further advantage is that the reflection scattering parameters are equivalent to the voltage and current reflection coefficients, since the normalising factors cancel out. Additionally, if the reference impedances are the same at each port (which is usually the case), the scattering parameters also provide the voltage and current transmission coefficients. These advantages may seem somewhat redundant for metallic waveguides where it is not convenient to think in terms of voltages and currents, but scattering parameters are widely used at lower frequencies where this is both possible and often done. In microwave engineering literature, the complex voltage reflection coefficient is usually denoted by ‘Γ’, and is equivalent to the $S_{nn}$ scattering parameter (where $n$ is the port number).

We can now summarise some important (and useful) points concerning scattering parameters:

- The $S_{mn}$ parameter is equivalent to the complex voltage reflection coefficient at Port $n$.
- The $S_{nm}$ parameter is equivalent to the voltage transmission coefficient from Port $n$ to Port $m$.
- Since the $a_n$ and $b_n$ signals have the dimensions of the square-root-of power, then $|a_n|^2$ and $|b_n|^2$ are the incident and emerging powers, so that $|S_{nn}|^2$ and $|S_{mn}|^2$ are the reflection and transmission coefficients for power.
- For a passive network (a network which does not add any power), the maximum magnitude for reflection coefficients is unity, i.e. $|S_{nn}| \leq 1$. Similarly, the magnitude for transmission coefficients cannot exceed unity, so we also have $|S_{mn}| \leq 1$.
- From energy conservation, the total scattered power in a passive network cannot exceed the incident power, and therefore $|S_{11}|^2 + |S_{21}|^2 \leq 1$.
- If the network does not absorb (dissipate, radiate, etc) any of the power, then the network is said to be ‘lossless’, and it follows that $|S_{11}|^2 + |S_{21}|^2 = 1$.

Display formats

Network analyzers display measurement results in the form of scattering parameters, usually as a function of frequency. They may also display other quantities which may be derived directly from the scattering parameters. It is worth looking carefully at the VNA’s display to gain familiarity with the various measurement formats available.

VNAs are generally designed with engineers in mind. Consequently, it is likely that the default measurement format is the $S_{11}$ parameter, displayed in decibels (dB). This is a logarithmic-magnitude format (‘Log Mag’), and since the decibel unit is defined for power ratios, this is simply,

---

7 Decibels are based on common (base 10) logarithms.
\[ S_{11}(\text{dB}) = 10\log_{10}|S_{11}|^2 \] (8.5)

or equivalently,

\[ S_{11}(\text{dB}) = 20\log_{10}|S_{11}| \] (8.6)

To interpret the display, the user needs to remember that a linear 1:1 reflection will be 0 dB, and a zero reflection will be \(-\infty\) dB. In practice, noise and other limitations mean that true zero reflections are not observed, and typically a nominal zero reflection will appear as a negative value in decibels somewhere between \(-50\) dB and \(-100\) dB. This corresponds to a reflected power ratio of between \(1 \times 10^{-5}\) and \(1 \times 10^{-10}\).

The VNA user may select other formats for the complex scattering parameters, including linear magnitude and phase (in degrees), and linear real and imaginary components. These are normally displayed on rectangular graphs, as a function of the test frequency.

The phase display may seem a little perplexing at first. It should be noted that the scale is usually in degrees, with positive angles above a ‘zero reference’ and negative angles below it. The key to interpreting this display format is to remember that the points \(+180^\circ\) and \(-180^\circ\) are coincident. Any measurement trace depicting a continuous, linear change of phase, as a function of frequency, will appear to have sharp discontinuities where the phase progresses between \(\pm 180^\circ\).

The linear complex scattering parameters may also be presented on a polar plot where the measurement trace indicates the scattering data as a function of frequency. The default scale is usually chosen such that the outer circumference corresponds to a unit circle. Thus, a maximum reflection coefficient for a passive DUT will produce a trace (as a function of frequency) that lies on this circle.

An impedance grid may be superimposed on the polar reflection scale, in which case we have the famous ‘Smith chart’ (popular with RF and microwave engineers) which is described in many texts [8]. The impedances are related to the characteristic impedance, \(Z_0\), of the ports (the measurement reference impedance). Where \(Z_0\) may be meaningfully assigned a value, the Smith chart impedances are de-normalised to give the actual impedance values associated with the network. In rectangular waveguide, where the characteristic impedance is not easy to define, it is common practice to assign the value of \(Z_0\) to 1. This simply means that any impedances displayed are really normalised impedances, relative to the waveguide impedance.

Figure 8.3 shows a selection of typical VNA measurements of a ‘one-port’ device (in this example, the device is a 94 GHz antenna connected by a short length of rectangular waveguide\(^8\)). Each display format is derived from the same primary measurement data, but presented in different ways. Notice that the phase of the reflection coefficient reduces (almost linearly) as the measurement frequency increases, and the transition from \(-180^\circ\) to \(+180^\circ\) does not indicate any significant discontinuity in the measured phase. Notice too, that the reflection coefficient \((S_{11})\) is a minimum at the frequency where the antenna is designed to radiate the energy away efficiently.

Most VNAs allow the user to add ‘markers’ to display the trace value at particular frequencies, and offer many other data processing/comparison functions. The measurement graphs in figure 8.3 have been kept relatively simple for clarity.

\(^{8}\) An antenna is a one-port network as far as the VNA measurement is concerned.
Figure 8.3: Selection of measurement formats available from a typical VNA: (a) linear magnitude; (b) log-magnitude (reflection coefficient in decibels); (c) phase of the reflection coefficient; (d) Smith chart.

Other parameters may be derived from the scattering parameter data, and are usually available to view directly on the VNA. For example, there is a one-to-one correspondence between the reflection coefficient ($S_{11}$, $S_{22}$) and the ‘standing wave ratio’ (SWR) [9], and most VNAs will display reflection data in this format.

One final data format available to the VNA user is worthy of mention here. The propagation delay [10] through a two-port network may also be determined from the rate of phase change as a function of frequency, for the steady-state transmission coefficients ($S_{21}$, $S_{12}$). The VNA computes the delay time by differentiating the phase measurement with respect to frequency.
8.1.3 VNA systems

It will be instructive to first consider the essential hardware features of the common VNAs used to measure scattering parameters from a few kilohertz to frequencies in the region of 50 GHz (an ‘RF/microwave VNA’). Figure 8.4 shows a block diagram of a typical two-port VNA.

![Block diagram of a typical two-port VNA](image)

Figure 8.4: Block diagram of a typical two-port vector network analyzer. The test signal proceeds through an electronically-controlled switch to a ‘signal separation’ device, and then to the VNA test port (measurement reference plane). Reflected signals, and signals transmitted through to the ‘other’ port, will travel back toward the switch/signal source. Ideally, both the switch and the signal source circuits are designed such that no re-reflection of these signals will occur.

This is a somewhat simplified view of the major components and system diagram of a two-port VNA, and of course, there are many variations found on commercial products. However, the basic operation is essentially the same. A signal source operating in a continuous-wave (CW) fashion is used to supply the test signal. This test signal is routed to one of the two ‘test ports’, usually by means of an electronically-controlled switch. This means that, ordinarily, only one test-port receives a

---

9 Some modern VNAs are ‘multiport’ instruments; our discussion here is confined to the more common two-port VNA.

10 The term ‘port’ is used for both the network or device being tested and also the VNA’s measurement terminals. The VNA has ‘test ports’, to which the ports of the DUT network are connected. As with the DUT, the VNA test ports can be realised with any electromagnetic guiding structure, (for example, co-axial cables or metallic waveguides) although clearly these need to be of the same type on both to facilitate connections. At microwave frequencies, ‘adapters’ are sometimes used to convert from one guiding structure to another.
stimulus signal at any given time (although the VNA will normally switch the test signal quite rapidly between the test-ports). This is consistent with the definition of scattering parameters mentioned earlier, where each parameter is obtained under the condition that only one incident (‘a’) signal is present at a given time.

Along the path between the signal source and the test port there are signal separation devices. These can take various forms, but most commonly they are implemented with components known as ‘directional couplers’. Effectively, these components separate the signals on the basis of direction of travel. Their role is quite simple: to ‘tap off’ a proportion of the outgoing (‘a’) and returning (‘b’) test signals to be measured within the VNA.

This architecture results in signals within the VNA which are proportionally related to the signals incident on and emerging from the DUT. At a given moment, there should only be one incident-related signal, (either a₁ or a₂). However, depending on the DUT, there may be two emerging-related signals (b₁ and b₂) present. Ratios of these internal signals form the basis of the scattering parameter measurements, but it should be noted that there will be inevitable amplitude and phase errors introduced by losses, internal mismatches, imperfect ‘directivity’ of the directional couplers, and arbitrary path lengths (phase shifts) in the VNA system. However, this is not a major problem, since these errors are systematic. Providing they can be determined (i.e. quantified) by a suitable calibration process, the measurements may be mathematically corrected to account for them (see the discussion below on calibration standards and methods).

Since the internal signals (related to the external ‘a’ and ‘b’ signals) are of the same high frequency as the measurement stimulus, they cannot be processed directly. The solution is to down-convert them to a much lower frequency. This is achieved by the use of frequency mixers which combine (i.e. multiply) the RF/microwave test signals with a second signal source (a local oscillator, ‘LO’) to produce an intermediate frequency (IF).¹¹ This process is known as heterodyning, and is essentially the same method used in good quality radio receivers to down-convert incoming signals to more convenient frequencies prior to de-modulation. Multiplying two sinusoids of different frequencies yields two new signals whose frequencies will be at the sum and difference of the frequencies of the original signals.¹² By careful choice of the LO frequency, and then selecting the difference frequency from the mixer output, the IF signal will be at a much lower frequency than that of the original measurement signal. Crucially, these IF signals preserve both the amplitude proportionality and the phase relationships of the original high-frequency signals. The VNA digitises the IF signals, and after appropriate manipulation (signal processing), displays the results.

Older network analyzer systems often had the measurement signal sources, directional coupler assemblies (known as the ‘test set’), IF/digitiser and display/processor units, as separate instruments connected by a daunting maze of cables. Such set-ups are still commonly found in laboratories, although with more modern VNAs, the entire system is housed within a single instrument ‘box’.

**Millimeter-wave systems**

For millimeter-wave, submillimeter-wave and THz VNAs, the basic architecture is essentially the same. The main difference the user will notice is that some of the key components are housed in separate modules (known as ‘extender modules’ or ‘extender heads’). These are usually intended to be used in conjunction with a standard microwave VNA, and thus a complete system typically consists

---

¹¹ Some VNA equipment have only one high-quality signal source, which is used to provide the measurement stimulus. The ‘local oscillator’ functionality is achieved through the use of ‘samplers’ which mix the test signals with harmonically-rich waveforms to obtain the necessary IF signals. Conversely, more modern VNAs, and especially systems which make use of ‘extender heads’ contain two high-quality signal sources. These may be standalone instruments in their own right, but are more often integrated into the VNA ‘box’.

¹² This easily follows from the trigonometric expansion of sinA sinB which yields terms of cos(A+B) and cos(A-B). Usually, only the difference term is useful.
of a microwave VNA, and one or two extender heads. Sometimes, the extender heads are designed and manufactured by a third-party company in close co-operation with the VNA manufacturer.

Figure 8.5 shows the typical hardware configuration for a VNA with extender modules. There may be additional external units to supply DC power to the extender heads which are not shown here. Also, some VNA systems may require an external controller to manage the exchange of RF/LO and IF signals between the VNA and the extender heads:

8.1.4 Extender heads

Figure 8.5 shows that the VNA system typically supplies the extender heads with a radio-frequency (RF) signal and a local oscillator (LO) signal. These signals are usually in the microwave portion of the frequency spectrum, typically a few Gigahertz.

For millimeter-wave, submillimeter-wave and THz systems, the actual stimulus test signal is generated within the extender head by a process of harmonic multiplying. The technique relies on driving a non-linear semiconductor device (usually a Schottky diode) with a moderately high-power RF signal. The non-linearity in the target device leads to the generation of many harmonics (integer multiples) of the RF signal. The power level at each harmonic frequency diminishes as the harmonic
number increases, and at the desired output frequency, the power is usually very low compared to the RF input signal.

Table 8.2 shows RF frequencies, harmonic numbers and test-port output powers for a variety of common measurement frequency ranges. These are typical values – individual VNA systems may have different values, depending on the manufacturer. Also, many manufacturers offer more than one extender head version for each frequency range in order to accommodate different VNA RF frequency capabilities.

<table>
<thead>
<tr>
<th>Desired test frequencies</th>
<th>Input RF frequency (GHz)</th>
<th>Harmonic number</th>
<th>Test-port output power</th>
</tr>
</thead>
<tbody>
<tr>
<td>110 GHz – 170 GHz</td>
<td>27 – 42</td>
<td>4</td>
<td>1 mW</td>
</tr>
<tr>
<td>140 GHz – 220 GHz</td>
<td>24 – 37</td>
<td>6</td>
<td>0.3 mW</td>
</tr>
<tr>
<td>220 GHz – 325 GHz</td>
<td>18 – 27</td>
<td>12</td>
<td>0.1 mW</td>
</tr>
<tr>
<td>325 GHz – 500 GHz</td>
<td>27 – 42</td>
<td>12</td>
<td>20 μW</td>
</tr>
<tr>
<td>500 GHz – 750 GHz</td>
<td>28 – 42</td>
<td>18</td>
<td>3 μW</td>
</tr>
<tr>
<td>750 GHz – 1.1 THz</td>
<td>21 – 31</td>
<td>36</td>
<td>0.3 μW</td>
</tr>
</tbody>
</table>

Table 8.2: Some typical RF frequencies used to drive the harmonic multipliers, and corresponding output power. Individual systems may follow the same operating principle, but use different harmonic numbers, and the actual output power may vary considerably.

The absolute value of the test port power is not important since scattering parameters are formed by ratios of incident and emerging signals. There is an implicit assumption here that the network being measured responds linearly to the test signal. If the network response is not linear, interpreting the measurement results requires additional care.

The extender heads also contain the directional signal separation components (couplers). The process of down-converting the coupled test signals to more convenient frequencies is similar to that employed in the microwave VNA. However, rather than mix the coupled signals with the LO signal directly, a harmonic of the LO signal is used (in a manner similar to the harmonic multiplying above). By driving a suitable non-linear semiconductor device with the LO signal, the test signal will mix (multiply) with the harmonics of the LO frequency. One of these harmonics will lead to an IF signal in a convenient frequency range which may be processed in the same manner as that used in the conventional microwave VNA. The LO mixer circuits are included as part of the extender head. In this manner, the millimeter-wave, submillimeter-wave or THz signals are all confined to the extender head and DUT.

The RF, LO and the down-converted signals are all connected to the VNA. The user does not normally need to be concerned with selecting the correct frequencies for the RF and LO signals; this is done automatically in the VNA, although the user may be required to enter the appropriate harmonic numbers in the VNA configuration.

**Transmission/reflection options**

A complete system for two-port measurements will include two identical extender heads. However, the extender heads are by far the most expensive part of the system, and most manufacturers offer a reduced-cost alternative. This is accomplished by reducing the complexity of one of the extender heads by omitting the components required to measure reflection. This leaves the user with a full

---

13 Apart from signal to noise ratio considerations, the absolute value of the test signal plays no part in the measurement, providing the DUT is linear, so that the scattering response is not a function of the power level.

14 Many modern VNA's will have this information either pre-loaded, or easily loaded via an instrument state file supplied on suitable media by the extender head manufacturer.
transmission/reflection head (‘a1’ and ‘b1’ signals) and a transmission-only head (just a ‘b2’ signal). The result is a system which can measure $S_{11}$ and $S_{21}$, but in order to measure the reverse parameters ($S_{22}$ and $S_{12}$) the DUT must be physically reversed. This is sometimes referred to as a ‘one-path, two-port’ system.

It is also entirely possible to have a system with only one ‘transmission/reflection’ extender head. Despite the terminology, such a system can only be used for reflection measurements! The extender head would be capable of measuring transmission if it were used in a system with a second extender head. The labelling really stems from the fact that, to measure reflection requires ‘a1’ and ‘b1’ signals, whereas to measure transmission requires only the ‘b’ signal at the second port.

![Diagram](image)

Figure 8.6: A lower-cost option is implemented when the second extender head is only capable of receiving signals. ‘Forward’ transmission and reflection can be measured, but the DUT must be reversed to measure ‘reverse’ reflection and transmission.

The VNA user needs to consider carefully the practical arrangement of the extender heads and the DUT. The block diagram of the connection scheme might seem straightforward, but the physical reality often presents some challenges. For millimeter-wave to THz frequencies, the extender heads are likely to connect to the DUT via rigid rectangular metallic waveguides. This implies that the orientation, spacing and alignment of the heads must be carefully thought-out. In the next section, we shall discuss some of the properties of metallic waveguides and their interconnections.

---

15 The reduced capability of the second extender head also has implications for the calibration process, and some methods of calibration are not available in these systems.
Figure 8.7: (a) VNA with extender heads. Flexible cables are used to connect the RF and LO harmonic multiplier/mixer drive signals, and return the IF output signals to the VNA. Also visible are DC power modules which are required to power the extender heads; (b) close-up of DUT between the extender head measurement reference planes. Here, the DUT is actually an additional length of precision waveguide.

8.1.5 Good practice tips

Tip 1: Plan the measurements from the perspective of physical connections before attempting to calibrate – or even switch on – the VNA. Check the orientation of the DUT waveguides, and verify that it is possible to actually make the necessary connections. The extender heads may need to be placed on suitable mounts, or perhaps rotated, or supported on stands. It is a good idea to reduce the mechanical strain on the waveguides as much as possible by supporting the weight of the extender heads, whilst allowing for fine movement of the heads for alignment. Some fairly elaborate looking air-cushion mounts and sliding rail systems have been used in some laboratories, although the authors have found that blocks of inexpensive packing foam can perform very well! The use of good quality laboratory jacks, possibly with precision adjustments (micrometers) should also be considered.

Tip 2: Avoid, as much as possible, moving the extender heads once a calibration and measurement activity has been commenced. Even small amounts of flexing on the cables which connect the RF, LO and down-converted test signals to the VNA can lead to phase-shifts, affecting the integrity of the calibration and ultimately reduced quality of measurements. Some movement is often necessary, but the user should try to minimise it as much as possible.

Tip 3: Use additional waveguide sections to extend the physical location of the test ports. This can not only make the physical connections more practical, it also serves to protect the original waveguide test ports from expensive damage. Adding lengths of additional waveguide will not significantly affect the measurement accuracy (the effects of the extra waveguide length are fully accounted for in the calibration process). However, at higher frequencies, some appreciable loss will be present in the waveguide; the main impact of this loss will be to reduce the available dynamic range of the measurement.

Tip 4: Measurements of waveguide devices with more than two ports are possible, even with a two-port VNA system. The scattering parameters for $n$ ports are acquired by measuring pairs of ports in turn. Under these circumstances, it is preferable to terminate ports that are not being measured with reflection-less (i.e. matched) terminations. This ensures that all incident signals are zero, apart from the port being stimulated by the VNA. Measuring $n$-port devices does present practical challenges though. The position and orientation of the extender heads, and the amount of head movement required just to make the necessary connections, is often problematic and requires considerable prior thought.
8.2 Metallic waveguides

8.2.1 Basic properties

Most of today’s VNAs that operate at frequencies above 110 GHz have test ports that are realised using rectangular metallic waveguide. These test ports are where the devices to be tested (e.g. electronic components, etc) are connected. On a two-port VNA, a one-port device (such as a termination) needs only to be connected to one of the two test ports. For two-port devices (such as attenuators, amplifiers, etc), these need to be connected to both of the VNA’s test ports.

Figure 8.8: A rectangular metallic waveguide of width, $a$, and height, $b$.

Rectangular metallic waveguide (or, ‘waveguide’, for short) has a long history of usage that dates back to the early-to-middle part of the 20th century. In those early days, waveguides were often used on high power systems (e.g. for radar applications) operating at microwave frequencies (i.e. at frequencies ranging from around 1 GHz to 30 GHz). The waveguides were found to be very suitable for building these types of systems because the waveguides exhibited relatively low electromagnetic losses (i.e. low attenuation). The waveguides were also able to handle the high power levels that were often used for the radar applications. Finally, it was relatively easy to join the waveguides together without causing any significant loss of signal (i.e. the junctions where the waveguides were joined generated very low levels of electromagnetic reflection).

A potential drawback with this type of waveguide is that, for most normal applications, it has an inherently narrow bandwidth (for mono-mode propagation\(^\text{16}\)). This was not a problem for the early microwave systems that tended to use only a very narrow range of frequencies (or sometimes only a single frequency was used). The size of waveguide was therefore chosen to accommodate the required operating frequency (or frequencies).

Figure 8.8 shows a diagram of a length of waveguide. The important mechanical dimensions of the waveguide are the width, $a$, and height, $b$. Most waveguides that are used these days have a width-to-height aspect ratio of 2:1 – i.e. the width of the waveguide aperture is twice the height of the aperture.

The bandwidth for mono-mode propagation for a given waveguide size is one octave. Below this frequency range, no electromagnetic energy can propagate in the waveguide. There is a well-defined

---

\(^{16}\) In general, these types of waveguide can operate over broad bandwidths. However, multiple electromagnetic modes can simultaneously be present over these broad bandwidths and this makes the performance of the waveguide difficult to predict. Conventional waveguide uses therefore usually restrict the bandwidth to ensure only a single mode of propagation is present – hence the name ‘mono-mode propagation’.
frequency below which the waveguide stops propagating energy, and this is called the ‘cut off frequency’. The cut off frequency, \( f_c \), is determined by the width (i.e. the broad wall dimension) of the waveguide, \( a \), as follows:

\[
f_c = \frac{c}{\sqrt{\varepsilon_r}} \times \frac{1}{2a}
\]  

(8.7)

where \( c \) is the speed of electromagnetic waves in vacuum (defined as 299 792 458 m/s) and \( \varepsilon_r \) is the relative permittivity of the material filling the waveguide. In most cases, the waveguide is filled with air (\( \varepsilon_r \approx 1 \)), and so the cut off frequency is given by:

\[
f_c = \frac{299792458}{2a}
\]  

(8.8)

Therefore, the frequency range for mono-mode propagation in waveguide is, in principle, from \( f_c \) to \( 2f_c \). However, in practice, the operational bandwidth is taken as less than an octave. This is to avoid problems that occur at frequencies close to \( f_c \) and \( 2f_c \). At frequencies close to \( f_c \), the dispersion (see subsection 8.2.2, below) and loss rapidly increases; at frequencies close to \( 2f_c \), there is an increased likelihood that higher-order modes will start to propagate and so mono-mode propagation inside the waveguide can no longer be assured.

To avoid these problems, a reduced operational bandwidth is usually given for a specific waveguide size. Generally, the suggested minimum frequency, \( f_{\text{min}} \), and maximum frequency, \( f_{\text{max}} \), for a given waveguide size (as defined by the waveguide aperture width, \( a \)) is given by:

\[
f_{\text{min}} \approx 1.25 \times f_c
\]  

(8.9)

\[
f_{\text{max}} \approx 1.9 \times f_c
\]  

(8.10)

The basic waveguide properties described above (i.e. low attenuation, high power handling capability, low reflection of waveguide connections) have enabled waveguide to be used successfully at microwave frequencies for many years. In more recent years, there has been an increased interest in the commercial use of frequencies in the millimeter-wave (30 GHz to 300 GHz) and submillimeter-wave (300 GHz to 3 THz) regions. At these higher frequencies, waveguide is still seen as an attractive transmission line for guiding electromagnetic signals. This is despite the attenuation being higher at these frequencies and the power handling capability being lower. In addition, the interfaces used to join two pieces of waveguide also become significantly more reflective – i.e. a significant amount of the signal being transmitted can be reflected by a slight change in cross-sectional geometry caused by two joining waveguide interfaces.

We will further discuss these ‘THz waveguide’ sizes, and their interfaces, in the sections that follow. But first we will describe another interesting feature affecting all waveguide sizes, and that is the property known as ‘dispersion’.

8.2.2 Dispersion

When a sinusoidal electromagnetic signal travels in free space, the free space wavelength, \( \lambda \), is related to the frequency, \( f \), as follows:

\[
\lambda = \frac{c}{f}
\]  

(8.11)

where, as before, \( c \) is the speed of electromagnetic waves in vacuum (defined as 299 792 458 m/s).
However, for a sinusoidal electromagnetic signal travelling inside a rectangular metallic waveguide, the wavelength (often called the ‘guide wavelength’) is considerably longer that the wavelength in free space. In fact the difference between the guide wavelength and the free space wavelength increases as the frequency decreases. When the frequency reaches the cut off frequency (below which no electromagnetic signal will propagate) the guide wavelength becomes infinite.

The guide wavelength, $\lambda_g$, is given by:

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (\lambda / \lambda_c)^2}}$$  \hspace{1cm} (8.12)

where $\lambda_c$ is the waveguide cut off wavelength, $\lambda_c = 2a$.

Equation 8.12 shows that when the free space wavelength reaches the cut off wavelength (i.e. when the frequency reaches the cut off frequency), $\lambda / \lambda_c = 1$ and so the denominator on the right hand side goes to zero and so $\lambda_g$ becomes infinite.

Figure 8.2 shows a plot of the guide wavelength and the free space wavelength versus frequency for a ‘standardized’ waveguide size that is used at frequencies from 500 GHz to 750 GHz. The difference between the free space and guide wavelengths is referred to as the dispersion. The guide wavelength is a non-linear function of frequency, and it can be seen that the lower frequency signals appear to propagate faster than the higher frequency signals. This effect is known as dispersion. Figure 8.9 clearly shows that the dispersion effect increases as the frequency decreases.

![Figure 8.9: Free space wavelength and Guide wavelength for a waveguide with width, $a = 0.38$ mm.](image)

The size of the waveguide used to calculate the wavelengths shown in figure 8.9 is a so-called ‘standardized’ waveguide size (see sub-section 8.2.3, below). This waveguide size has an aperture width of, $a = 0.38$ mm. The cut off frequency for this waveguide (using equation 8.8) is therefore $f_c \approx 394$ GHz and the recommended minimum and maximum frequencies (using equations 8.9 and 8.10) are $f_{\text{min}} \approx 493$ GHz and $f_{\text{max}} \approx 749$ GHz. In practice, the minimum and maximum suggested
operating frequencies for this particular waveguide size are usually rounded to be 500 GHz and 750 GHz, respectively.

Figure 8.9 shows that the dispersion increases rapidly below the recommended minimum frequency of 500 GHz and this is one of the reasons why it is not advisable to use this waveguide size below approximately 500 GHz. Although the dispersion is much lower at frequencies close to 800 GHz, the recommended maximum frequency is set to 750 GHz to ensure that the waveguide will only accommodate mono-mode propagation. (Note that the calculated frequency for the onset of higher order modes for this waveguide size is \(2f_c \approx 788\) GHz.)

The dispersion characteristics for other waveguide sizes show similar behaviour to those shown in figure 8.9. So, in general, for any waveguide size, the suggested useful frequency range can be established as typically 1.25\(f_c\) to 1.9\(f_c\), where \(f_c\) is determined from the width of the waveguide aperture (using equation 8.8).

8.2.3 Standardized sizes and frequency ranges

The above general properties of waveguides imply that any size of waveguide can be used – the choice of waveguide size being governed by which frequencies are of interest. However, in practice, there are many advantages with using a relatively small set of waveguides sizes. One such advantage is that these sizes could be used by others and so it makes it easy to share and exchange components due to the compatibility of their sizes.

This has been recognised for many years by users of waveguide at microwave and millimeter-wave frequencies. In fact agreed ‘standardised’ sizes of waveguides have been in use at frequencies up to 330 GHz for many years [11, 12]. In recent years, a similar ‘standardization’ activity has taken place addressing waveguide sizes suitable for the submillimeter-wave / terahertz frequency ranges (i.e. waveguides sizes used for frequencies above 330 GHz). This standardization activity has been led by the IEEE and has resulted in a new standard: IEEE Std 1785.1-2012 [13]. This standard provides sizes that are based on the same approach that was used to define the sizes used at the lower microwave and millimeter-wave frequencies [11, 12]. This has resulted in two contiguous interleaved series (containing no gaps or overlaps in the frequencies covered by each series). In addition, the new IEEE standard defines the waveguide widths using metric units (i.e. \(\mu m\)).

Table 8.3 shows the new standardized waveguide sizes and frequency ranges, where, \(a\) and \(b\) are the waveguide aperture width and height, respectively; \(f_{\text{min}}\) and \(f_{\text{max}}\) are the suggested minimum and maximum frequencies, respectively; and \(f_c\) is the cut-off frequency. The names of the waveguide sizes (WM-2540, WM-2032, etc) follow a new naming convention that has been developed for these waveguide bands. Since the sizes are defined in terms of metric units, the letters WM are used to indicate that the size refers to Waveguide using Metric dimensions. These letters are then followed by a number that indicates the size (in \(\mu m\)) of the waveguide width (i.e. broad wall) dimension.

The two contiguous interleaved series in Table 8.3 are as follows: (i) WM-710, WM-470, WM-310, etc; and (ii) WM-570, WM-380, WM250, etc. Both series can be extended to higher frequencies (if needed), as follows:

1. Use the waveguide sizes that are unshaded in Table 8.3;
2. Divide the mechanical dimensions by 10;
3. Multiple the frequencies by 10;
4. Rename the waveguide accordingly.
<table>
<thead>
<tr>
<th>Waveguide name</th>
<th>(a) (μm)</th>
<th>(b) (μm)</th>
<th>(f_c) (GHz)</th>
<th>(f_{\text{min}}) (GHz)</th>
<th>(f_{\text{max}}) (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM-2540</td>
<td>2540</td>
<td>1270</td>
<td>59.014</td>
<td>75</td>
<td>110</td>
</tr>
<tr>
<td>WM-2032</td>
<td>2032</td>
<td>1016</td>
<td>73.767</td>
<td>90</td>
<td>140</td>
</tr>
<tr>
<td>WM-1651</td>
<td>1651</td>
<td>825.5</td>
<td>90.790</td>
<td>110</td>
<td>170</td>
</tr>
<tr>
<td>WM-1295</td>
<td>1295</td>
<td>647.5</td>
<td>115.75</td>
<td>140</td>
<td>220</td>
</tr>
<tr>
<td>WM-1092</td>
<td>1092</td>
<td>546</td>
<td>137.27</td>
<td>170</td>
<td>260</td>
</tr>
<tr>
<td>WM-864</td>
<td>864</td>
<td>432</td>
<td>173.49</td>
<td>220</td>
<td>330</td>
</tr>
<tr>
<td>WM-710</td>
<td>710</td>
<td>355</td>
<td>211.12</td>
<td>260</td>
<td>400</td>
</tr>
<tr>
<td>WM-570</td>
<td>570</td>
<td>285</td>
<td>262.97</td>
<td>330</td>
<td>500</td>
</tr>
<tr>
<td>WM-470</td>
<td>470</td>
<td>235</td>
<td>318.93</td>
<td>400</td>
<td>600</td>
</tr>
<tr>
<td>WM-380</td>
<td>380</td>
<td>190</td>
<td>394.46</td>
<td>500</td>
<td>750</td>
</tr>
<tr>
<td>WM-310</td>
<td>310</td>
<td>155</td>
<td>483.53</td>
<td>600</td>
<td>900</td>
</tr>
<tr>
<td>WM-250</td>
<td>250</td>
<td>125</td>
<td>599.58</td>
<td>750</td>
<td>1100</td>
</tr>
<tr>
<td>WM-200</td>
<td>200</td>
<td>100</td>
<td>749.48</td>
<td>900</td>
<td>1400</td>
</tr>
<tr>
<td>WM-164</td>
<td>164</td>
<td>82</td>
<td>913.99</td>
<td>1100</td>
<td>1700</td>
</tr>
<tr>
<td>WM-130</td>
<td>130</td>
<td>65</td>
<td>1153.0</td>
<td>1400</td>
<td>2200</td>
</tr>
<tr>
<td>WM-106</td>
<td>106</td>
<td>53</td>
<td>1414.1</td>
<td>1700</td>
<td>2600</td>
</tr>
<tr>
<td>WM-86</td>
<td>86</td>
<td>43</td>
<td>1743.0</td>
<td>2200</td>
<td>3300</td>
</tr>
</tbody>
</table>

Table 8.3: Frequency bands and waveguide dimensions used in IEEE Std 1785.1-2012 [13].

For example, the next two sizes in the series (derived from WM-710 and 570) are shown in Table 8.4.

<table>
<thead>
<tr>
<th>Name</th>
<th>(a) (μm)</th>
<th>(b) (μm)</th>
<th>(f_c) (GHz)</th>
<th>(f_{\text{min}}) (GHz)</th>
<th>(f_{\text{max}}) (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM-71</td>
<td>71</td>
<td>35.5</td>
<td>2111.2</td>
<td>2600</td>
<td>4000</td>
</tr>
<tr>
<td>WM-57</td>
<td>57</td>
<td>28.5</td>
<td>2629.7</td>
<td>3300</td>
<td>5000</td>
</tr>
</tbody>
</table>

Table 8.4: Extended frequency bands and waveguide dimensions using IEEE Std 1785.1-2012 [13].

The shaded region in Table 8.3 indicates where the new IEEE standard overlaps with earlier standards, e.g. the WR sizes given in MIL standard MIL-DTL-85/3C [11]. In this overlap region, the WM sizes have been set to be the same as the earlier WR sizes (that were specified using ‘Imperial’ units, i.e. inch and mil). This equivalence is shown in Table 8.5.

<table>
<thead>
<tr>
<th>MIL name</th>
<th>New IEEE Name</th>
<th>(f_{\text{min}}) (GHz)</th>
<th>(f_{\text{max}}) (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR-10</td>
<td>WM-2540</td>
<td>75</td>
<td>110</td>
</tr>
<tr>
<td>WR-08</td>
<td>WM-2032</td>
<td>90</td>
<td>140</td>
</tr>
<tr>
<td>WR-06</td>
<td>WM-1651</td>
<td>110</td>
<td>170</td>
</tr>
<tr>
<td>WR-05</td>
<td>WM-1295</td>
<td>140</td>
<td>220</td>
</tr>
<tr>
<td>WR-04</td>
<td>WM-1092</td>
<td>170</td>
<td>260</td>
</tr>
<tr>
<td>WR-03</td>
<td>WM-864</td>
<td>220</td>
<td>330</td>
</tr>
</tbody>
</table>


Finally, Table 8.6 shows a comparison between the new IEEE waveguides names and names resulting from a previous attempt to extend the MIL standard naming convention to these submillimeter-wave frequencies. These ‘extended MIL’ names (WR-2.8, WR-2.2, etc) [14] are still often found describing waveguides that are used as these frequencies.
<table>
<thead>
<tr>
<th>'Extended MIL' name</th>
<th>New IEEE Name</th>
<th>( f_{\text{min}} ) (GHz)</th>
<th>( f_{\text{max}} ) (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WR-2.8</td>
<td>WM-710</td>
<td>260</td>
<td>400</td>
</tr>
<tr>
<td>WR-2.2</td>
<td>WM-570</td>
<td>330</td>
<td>500</td>
</tr>
<tr>
<td>WR-1.9</td>
<td>WM-470</td>
<td>400</td>
<td>600</td>
</tr>
<tr>
<td>WR-1.5</td>
<td>WM-380</td>
<td>500</td>
<td>750</td>
</tr>
<tr>
<td>WR-1.2</td>
<td>WM-310</td>
<td>600</td>
<td>900</td>
</tr>
<tr>
<td>WR-1.0</td>
<td>WM-250</td>
<td>750</td>
<td>1100</td>
</tr>
</tbody>
</table>


8.2.4 Flanges / Interfaces

Having discussed the sizes of waveguide that are used for propagating the electromagnetic signals, an equally important aspect is the method used to join together (or interconnect) sections of waveguide. As with the waveguide sizes discussed in the previous section, there is a long history of usage of interconnect mechanisms for waveguides used at microwave and millimeter-wave frequencies. The waveguide interconnect mechanism that enables sections of waveguide to be joined together is often called the ‘flange’ or ‘interface’.

A variety of different shapes and sizes of flange are used at microwave frequencies. However, for millimeter-wave frequencies, there is one flange design that is used much more than any other. This is the so-called ‘UG-387’ design, as shown in figure 8.10. This design has been in existence for many years and has been standardized in the MIL-DTL-3922/67E standard [15]. At the centre of the waveguide flange is the waveguide aperture. The waveguide aperture shown in figure 8.3 is relatively large (3.8 mm × 1.9 mm) and so is used at low millimeter-wave frequencies (50 GHz to 75 GHz). The flange also features four relatively large, threaded, holes (shown at the North, South, East and West positions in figure 8.10). Screws are inserted into these holes to enable two flanges to be bolted together and tightened. The method for aligning two flanges that are being connected together is to use alignment holes and alignment pins (as identified in figure 8.10). These alignment pins are usually permanently fitted to the waveguide flange. The two alignment pins on one of the waveguide flanges fit into the two alignment holes on the other waveguide flange, and vice versa. So, a total of four alignment pins (two per flange) fit into a total of four alignment holes (two per flange). The positions and tolerances of these holes and pins translate into providing the alignment for the apertures of the two waveguides being joined together.

Figure 8.10: A conventional UG-387 flange, used extensively at millimeter-wave frequencies.

However, the alignment achieved using conventional UG-387 flanges is often not sufficient for many present day applications. This is particularly the case for applications at high millimeter-wave frequencies (i.e. above 110 GHz). This is because, as frequency increases, the required size of the waveguide aperture decreases (as discussed in sub-section 8.2.3). Therefore, any tolerance and
positional error in the alignment pins and holes on the waveguide flanges will have a proportionally larger impact on the achieved alignment of the waveguide apertures, when these apertures are of a smaller size (i.e. for higher frequency applications).

The above problem has led to the introduction of a number of ‘precision’ UG-387 flange types. The most popular of these precision UG-387 flanges is shown in figure 8.11. The main design change with this flange, compared with the conventional UG-387 flange, is the introduction of two additional alignment holes, immediately above and below the waveguide aperture. These inner holes are machined to a tighter tolerance than the outer pins and holes found on the conventional UG-387 flange. In addition, separate detachable dowel pins are inserted into these inner holes during connection, resulting in a significant improvement in the alignment of the waveguide apertures.

The main design change with this flange, compared with the conventional UG-387 flange, is the introduction of two additional alignment holes, immediately above and below the waveguide aperture. These inner holes are machined to a tighter tolerance than the outer pins and holes found on the conventional UG-387 flange. In addition, separate detachable dowel pins are inserted into these inner holes during connection, resulting in a significant improvement in the alignment of the waveguide apertures.

![Figure 8.11: 'Precision' UG-387 flange, showing the two additional inner dowel holes situated immediately above and below the waveguide aperture.](image)

However, as the required frequency extends still further into the submillimeter-wave and THz frequency ranges, the ‘precision’ UG-387 is no longer adequate (in terms of providing a waveguide connection with relatively low reflection loss) for many applications. This has led to new flange designs being developed specifically for applications at these frequencies. As with the recent work on standardizing waveguide aperture sizes (in IEEE Std 1785.1-2012 [13]) above 110 GHz, work is now underway to standardize these new flange designs so that the flanges can be manufactured to agreed dimensions and tolerances. The three new designs that are being standardized are described below. All three designs can provide acceptable connection performance for many applications at submillimeter-wave and terahertz frequencies. Only brief descriptions are given here, as these designs are still in the process of being finalized. Once the designs have been finalized, they will be published in IEEE Std 1785.2 [16].

**Precision Dowel design**
To the naked eye, a Precision Dowel flange looks the same as the ‘precision’ UG-387 design (Figure 8.11). There are two design features that distinguish the Precision Dowel design from the ‘precision’ UG-387 design:

1. The diameter, position and tolerance of the two inner dowel holes are more tightly defined
2. The two dowel pins that fit into the inner dowel holes are of different diameters:
   - The larger diameter dowel pin provides planar alignment of the waveguide apertures
   - The smaller diameter dowel pin provides angular alignment of the waveguide apertures

   The two dowel pins are marked so that they can be clearly identified.
**Ring-Centered design**
As with the Precision Dowel flange, the Ring-Centered flange also looks (to the naked eye) the same as the ‘precision’ UG-387 design (as shown in figure 8.11). The main design feature with the ring-centered design is the use of a detachable precision centering ring to align the waveguide flanges (rather than the inner dowel pins that are used with the Precision Dowel design). Figure 8.12 shows photos of the ring-centered design, before and after inserting the centering ring.

![Figure 8.12: ‘Ring-Centered’ flange design, showing (a) flange before attaching centering ring, and (b) flange with centering ring attached.](image)

**Plug and Jack design**
This design comes as two parts: a Plug flange and a Jack flange. (Other names that could be used to describe the two parts are “plug and socket” or “male and female”.) As with the Precision Dowel flange and the Ring-Centered flange, the Plug flange also looks (to the naked eye) the same as the ‘precision’ UG-387 design (shown in figure 8.11). The Plug and Jack flanges are shown in figure 8.13. A connection can only be made between two waveguides when one waveguide has a Plug flange and the other has a Jack flange – a Plug flange must not be connected to another Plug flange, and, a Jack flange must not be connected to another Jack flange. Alignment of two waveguides, connected using the Plug and Jack flanges, is achieved by the precision fit between the concentric mating mechanism of the Plug and Jack sections of the two flanges.

![Figure 8.13: ‘Plug and Jack’ flange design, showing the Plug version (on the left) and the Jack version (on the right).](image)

All three flange types – Precision Dowel, Ring-Centered, and Plug sand Jack – are mechanically compatible with each other (and compatible with the earlier UG-387 and ‘precision’ UG-387 designs). However, to achieve the best electrical performance for a given flange connection, it is
necessary to only join together similar flange types – i.e. two Precision Dowel flanges, or, two Ring-Centered flanges, or, a Plug and a Jack flange. Typical performance values (in terms of worst-case reflection coefficient), for these flange type pairings used in three different waveguide sizes, are shown in Table 8.7.

<table>
<thead>
<tr>
<th>Waveguide Name</th>
<th>Minimum frequency (GHz)</th>
<th>Maximum frequency (GHz)</th>
<th>Worst-case reflection coefficient (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM-570</td>
<td>330</td>
<td>500</td>
<td>-32</td>
</tr>
<tr>
<td>WM-380</td>
<td>500</td>
<td>750</td>
<td>-26</td>
</tr>
<tr>
<td>WM-250</td>
<td>750</td>
<td>1100</td>
<td>-19</td>
</tr>
</tbody>
</table>

Table 8.7: Worst-case reflection coefficients for the three new flange types in selected waveguide sizes.

8.2.5 Good practice tips

Although rectangular metallic waveguide is generally a very mechanically robust transmission line type, there are some good practice tips that should be followed in order to ensure that waveguide components remain in good working order and that waveguide connections are reliable and of a good quality.

**Tip 1**: When connecting two waveguides fitted with flanges (such as the types described in subsection 8.2.4) make sure the alignment mechanisms are not subjected to any mechanical stress as the flange faces are brought together. For example, for flanges that use dowel pins for alignment, make sure the dowel pins move easily into the alignment holes and that they are not causing any significant friction. The flanges should first be brought together using the alignment pins before any screws are used to tighten the connection.

**Tip 2**: Use a torque driver to help ensure that screws are tightened to a uniform torque. Although different waveguide manufacturers are likely to specify different values of torque for the waveguide connections, it is a good idea to make connections with repeatable torque (i.e. using a torque driver). Figure 8.14 (a) shows a typical torque driver that is used for tightening waveguide flange screws to a specified torque. Figure 8.14 (b) shows typical screws that are used for making waveguide connections. Note that the screw has a hexagonal hole into which the torque driver fits. The torque driver has a rounded end so that it can fit into the screw’s hexagonal hole from a variety of different angles. This can be very useful if the flange screws are not easily accessible, due to other obstacles that might be in the way.

![Figure 8.14: (a) a typical torque driver used to tighten waveguide flange screws; (b) typical waveguide flange screws (with hexagonal screw holes).](image-url)
Tip 3: When tightening waveguide flange screws, tighten each screw gradually, and in stages. One method is to gradually tighten screws that are on opposite sides of the flange face – for example, tighten the screws at north and south, then tighten the screws at east and west, then repeat the process until all screws have reached the required tightness. This ensures even pressure is applied to the two waveguide flange mating faces as they are brought together by the tightening of the screws. Never just tighten one screw fully before some tightening of the other screws has taken place – this can cause the flange faces to ‘cock’ and so the waveguides are not joined together very evenly.

Tip 4: Support waveguide components, using your hand, as they are connected together, both during the alignment and screw-tightening stages. This prevents the dowel pins and/or screws from artificially binding during the connection process.

Tip 5: Use x-, y-, z-position adjusters (e.g. controlled by micrometers) for aligning bulky items, such as the VNA extender heads or heavy components. As with Tip 4, this prevents the flange dowel pins and screws from binding artificially during connection.

Tip 6: Avoid excessive handling of certain items as this can lead to thermal changes. Since waveguide is made of metal, waveguide devices make very good conductors of heat. For example, holding a waveguide load can cause heat to be conducted inside the waveguide and onto the resistive element used to form the load. An increase in temperature can cause the reflection of the load to change significantly.

8.3 Calibration Standards and Methods

8.3.1 Calibration: general principles

In the foregoing introduction to VNAs, we stated that any measurement of incident and emerging signals will be subject to inevitable systematic errors. For example, the devices used to separate the incident and emerging signals may not be perfect, such that the observed ratio of ‘a’ and ‘b’ signals will be in error (a ‘directivity’ error). Some of the signals which are reflected from the DUT back to the originating test-port, or by transmission through the DUT to a different test-port, may be re-reflected from the VNA (a ‘source-match’ error). There will also be attenuation and phase-shifts in the various signal paths within the VNA. To further complicate matters, these errors will be different for each frequency across the range of interest!

Fortunately, all of these systematic errors can be grouped together into a relatively simple ‘error model’. A brief survey of literature on VNA calibration will probably give the impression that the topic concerns a large amount of tedious mathematics [17]. Whilst this may be true to an extent, the underlying principles are quite straightforward. A simple illustration may help:

Consider a hand-held resistance meter of the type found in many laboratories. This sort of instrument is typically supplied with a pair of test leads, through which a small current is passed. The resistance of whatever device is placed between the test leads is determined by measuring the current that flows when a small test voltage is applied, using the well-known relationship of Ohm’s law. However, if the test leads themselves have some resistance, the meter will measure the sum of the test lead resistances plus the ‘DUT’ resistance. For example, if the leads have a resistance of 1 Ohm each, and the DUT is a 10 Ohm resistance, the meter will display a reading of 12 Ohms (an error of 20%!). The solution is quite straightforward: If we first measure a resistance (a ‘calibration standard’) for which we already know (with confidence) what the correct measurement result should be, we can easily determine the combined test lead resistance. An ideal choice of calibration standard in this example would be a short-circuit because it is easy to implement, simply by connecting the test leads together. We can
also have a high level of confidence in what the ‘correct’ result should be, i.e. zero Ohms. Any non-zero reading obtained from the calibration step may be subtracted from all subsequent measurements in order to correct the results.

In the case of the VNA, there are many systematic errors, but these may also be combined into a simplified error model. For a one-port VNA measurement, this model takes the form of an additional two-port network (sometimes called an ‘error-box’ network) which ‘intervenes’ between the desired measurement reference plane, and the actual measurements made within the VNA. Such a model represents all the possible errors, including arbitrary amplitude and phase errors, imperfect coupler directivity, and internal mismatches.

Figure 8.15 illustrates this concept:

![Diagram of error-box network](image)

Figure 8.15: For a one-port measurement, the systematic errors may be combined into an additional ‘error-box’, a two-port network which comes between the DUT and the actual measurement. This additional network accounts for all the systematic errors – imperfect directional signal separation (directivity), imperfect proportionality (tracking) and unwanted re-reflections from the VNA (source-match).

This intervening two-port ‘network’ can be fully described by a set of scattering parameters in the same way that any real two-port network can be. If these ‘error’ scattering parameters can be found, then it is possible to express the unknown one-port DUT scattering parameter (i.e. the reflection coefficient, \(S_{11}\)) in terms of the error parameters and the internal, uncorrected measurement value.

If \(S'_{11}, S'_{21}, S'_{12}\) and \(S'_{22}\) are the scattering parameters for the intervening ‘error network’, \(S_{11}\) is the true reflection coefficient of the DUT, and \(w\) is the ‘uncorrected’ internal measurement, then it is easy to show from equation 8.2 and 8.3 that,

\[
w = S'_{11} + \frac{S'_{21}S'_{12}S_{11}}{1 - S'_{22}}
\]

(8.13)

from which,

\[
S_{11} = \frac{w - S'_{11}}{S'_{22}(w - S'_{11}) + S'_{21}S'_{12}}
\]

(8.14)

Thus, if the error coefficients (\(S'_{11}, S'_{21}, S'_{12}\) and \(S'_{22}\)) are first determined, the VNA can mathematically correct the subsequent measurements.\(^{17}\) Although there are four error parameters, it may be seen from

\(^{17}\) It is worth noting that equations 6.13 and 6.14 appear in the literature in a variety of different forms; however, the underlying equations are exactly the same.
equation 8.13 and 8.14 that it is unnecessary to determine $S'_{21}$ and $S'_{12}$ individually, and hence there are only three independent parameters to find.

In the one-port error model, the term $S'_{11}$ corresponds to a directivity error (which should ideally be zero), and the term $S'_{22}$ corresponds to a mismatched source impedance (which again, should ideally be zero). The product of $S'_{21}S'_{12}$ corresponds to the various attenuation and phase shifts in the signal paths (sometimes called ‘tracking errors’); ideally, $S'_{21}S'_{12}$ should be equal to exactly one (i.e. no attenuation or phase shift). In practice, none of these ideal characteristics are found in real VNA hardware, and thus a calibration process is an essential step to achieving good measurements. In a manner similar to the resistance meter example above, measuring three calibration standards (with known reflection properties) will permit the error coefficients to be found by solving a system of three linear equations, based on equation 8.13.

The reader may wonder why the process of calibration, i.e. finding the error coefficients, cannot be undertaken once by the manufacturer, with the values stored internally in the VNA. The primary reason for not doing this is that the user is free to extend the signal path (by means of additional waveguides, etc) to any arbitrary location. Calibrating ‘locally’ in this way ensures that all of the systematic errors, including the extended signal path, are included in the error coefficients. Also, the calibration process effectively determines the reference plane (see section 8.1.1 above) for the measurement. As a result of adopting this philosophy, a locally performed calibration is essential in order to obtain accurate, and indeed meaningful, VNA measurements.

Two-port errors
The error model for two-port measurements may be developed from the one-port case. However, additional error coefficients are included such that the two-port error model comprises more than just a pair of one-port models. For example, the tracking error for a transmission measurement (say, $S_{21}$) will be different from the tracking error for a reflection measurement. There may also be a signal leakage path between the two test-ports (‘crosstalk’) which will affect measurements of high-isolation networks (DUTs where the true value of $S_{21}$ or $S_{12}$ is very small). Calibrating for two-port measurements will generally require both reflection (one-port) and transmission (two-port) calibration ‘standards’.

From the VNA user’s perspective, the process of actually performing a calibration is not particularly onerous. With the exception of a couple of small steps discussed shortly, the VNA user is not required to undertake any of the tedious mathematical calculations, which are always accomplished automatically within the VNA firmware. All that is required of the user is to physically connect the appropriate calibration standards, in turn, to the VNA test-ports.

Other sources of error
Before we consider practical calibration techniques, it is worth noting some limitations of the calibration process. In addition to systematic errors, measurements are also subject to random errors (such as noise in the measurement circuits, limited repeatability of waveguide connections, etc). Calibration procedures cannot provide any correction for random errors.

However, the VNA user can take steps to reduce the impact of random error sources. For example, measurement data will be subject to noise (‘trace noise’)$^{18}$. This can be reduced by averaging successive measurements, and most VNAs allow the user to select an averaging factor. Large values for the averaging factor will inevitably slow down the measurement time, although moderate

---

$^18$ Noise is unavoidably added to the measurement data, mainly in the receiver circuitry. It appears at all measurement signal levels as ‘trace noise’. The noise level also determines the minimum signal that can be detected, which governs the dynamic range of the measurement.
averaging will have little detrimental effect on performance.\textsuperscript{19} Whilst averaging can reduce the effect of noise, the amount of noise entering the measurement circuits can also be reduced by limiting the bandwidth of the down-converted IF signal. For optimal reduction of noise, both averaging and IF bandwidth reduction can be used.

Other random errors (such as flange connection repeatability) cannot be easily reduced, but some knowledge of their statistical impact can help to quantify the typical uncertainty of any calibrated measurements (see the discussion below in section 8.4).

Dealing with systematic errors through calibration assumes that the systematic errors themselves are constant and do not change with time, or at least, not within the time it takes to perform the desired measurements. Unfortunately, real measurements are beset by drift in the systematic errors. There are two main causes of drift, both of which can be mitigated with careful practice.

The first cause of drift is physical movement of any part of the signal path. The components within the VNA itself, and also within the millimeter-wave extender heads, are unlikely to move. The ‘weak link’ in terms of physical movement is the flexible cables between the VNA and the extender heads. Movement of these cables leads to small changes in attenuation and phase-shift, and since these parameters are part of the systematic errors, any changes which take place after calibration will degrade the measurement accuracy. It is therefore important to minimise the amount of movement of the extender heads once the calibration process has been undertaken.

A second cause of drift is thermal fluctuation of either the measurement signal path or the electronic circuits within the VNA and the extender heads. Changes in temperature will produce small changes in physical dimensions through linear expansion. These can be surprisingly significant at submillimeter-wave and THz frequencies. The effects of thermal drift can be reduced by allowing the VNA equipment to thermally stabilise before use. In practice, this can mean having the equipment switched on for several hours before calibration and measurements take place.

It is also important to maintain a stable ambient temperature. The flexible cables used to connect the extender heads to the VNA, and the waveguides used to connect the extender heads to the DUT, are more exposed to ambient thermal fluctuation than other parts of the signal path. Ideally, measurements should take place in a temperature-controlled laboratory to minimise the possibility of thermal changes to these parts of the signal path.

8.3.2 Types of calibration standard

Since the calibration process has a significant influence on the integrity of VNA measurements, the choice of calibration standards is very important. The quality of the final measurements will strongly depend on how accurately the complex (magnitude and phase) reflection and transmission properties of the calibration standards are known. In theory, the actual values for the reflection or transmission properties of calibration standards are arbitrary, providing we know their characteristics, and providing they are sufficiently different from each other. (The latter requirement arises from the fact that the solution to the error correction equations will be ill-conditioned if the values are too closely spaced.)

\textsuperscript{19} Some VNAs permit the user to choose between ‘trace averaging’ and ‘point averaging’. Trace averaging will calculate and display a new trace average after measuring each complete set of frequency points in the current range. Changing the DUT leads to the data trace appearing to ‘settle’ (slowly) toward the new measurement, as the influence of the ‘old’ data is progressively reduced through the averaging process. This behaviour can be undesirable, especially with large averaging factors. In point averaging, the VNA measures each frequency point multiple times (according to the averaging factor) and updates the trace data at that point, before moving to the next frequency point.
A VNA equipped with extender heads is likely to be supplied with a ‘calibration kit’, containing a selection of calibration standards. Clearly, the VNA must ‘know’ the reflection and transmission properties of the standards in order to solve for the error coefficients. In most cases, it is not necessary for the user to enter the characteristics of individual calibration standards into the VNA. This data is either pre-loaded into the VNA (for commonly used commercial calibration kits) or it is supplied on a USB memory device (or other medium) from the calibration kit manufacturer. Occasionally, it may be necessary to manually enter the properties associated with user-fabricated calibration kits, and some types of TRL kit (discussed below).

To understand why certain items are chosen as calibration standards, we must consider how easy (or difficult!) it is to be confident of the associated reflection or transmission properties. In practice, this is more readily achieved with some devices than others. For example, in metallic waveguide, we can predict (with reasonable confidence) the electrical properties of a short-circuit from basic physical principles. (A short-circuit made from a conducting metallic plate, placed across the waveguide aperture, constitutes an electric ‘wall’, from which a total reflection must be produced.)

‘Calculable’ reflection and transmission standards
The complex reflection coefficient for any waveguide termination may be calculated by recognising that the amount of reflection will depend on the relationship between the termination impedance and the characteristic impedance of the waveguide structure.

If $Z_T$ is the impedance of the termination, and $Z_0$ is the characteristic impedance of the waveguide, then the reflection coefficient at the termination, $\Gamma_T$, (i.e. the scattering parameter, $S_{11}$) is given by:

$$\Gamma_T = S_{11} = \frac{Z_T - Z_0}{Z_T + Z_0} \quad (8.15)$$

The impedance of the short-circuit is zero, because the electric field tangential to a perfect conductor must be zero. Consequently, regardless of how the waveguide impedance is defined, the value for the reflection coefficient for a short circuit termination is $-1$ (unity magnitude, with relative phase of the reflected and incident waves of 180°). Thus a short-circuit makes a reliable and easy-to-implement calibration standard. In a waveguide calibration kit, the short-circuit is usually supplied as a waveguide flange without any waveguide ‘aperture’, plus some convenient means to hold the device (see photo in figure 8.16). This is also known as a ‘flush’ short-circuit (or just ‘flush short’).

A short length of waveguide (of the same aperture as that used for the test ports) will produce an electrical delay (i.e. a phase-shift, according to the guided wavelength). By application of equations similar to those described in the previous section (equation 8.12) this phase shift can be calculated with a high degree of confidence. Thus, a short length of waveguide is also a useful calibration standard. Calibration kits often include one or more short lengths of waveguide, often called ‘shims’, or ‘offset shims’.

A short-circuit and a short length of waveguide may also be combined to form an ‘offset short-circuit’. Again, this device will have calculable electrical properties (a total reflection, with a modified phase-shift).

Some waveguide calibration kits have offset short-circuits manufactured as a single device. These standards may look similar to the flush short circuits, but there is a waveguide aperture on the surface of the flange face, leading to a short, rectangular recess which implements the offset short-circuit.

---

20 The phase shift will be different for each measurement frequency, and must therefore be calculated (by the VNA) at each frequency point for the calibration/measurement.
Open-ended (i.e. ‘open-circuited’) waveguides are not usually used for calibration. This is because a portion of the signal energy will radiate away from the open end of the guide. Some of this radiated energy may subsequently be reflected back toward the waveguide from nearby objects, in a manner which cannot be predicted. This makes it difficult to have confidence in the expected reflection response, and thus unsuited to a calibration standard. However, some recent experimental calibration methods have been proposed that do utilise an open-ended waveguide as a calibration ‘standard’ [18]. These techniques have not yet been widely adopted, mainly due to the practical requirement to ensure a reflection-less environment in front of the open waveguide. They do have one advantage, which is that problems with flange misalignments which affect offset short-circuits do not arise.

‘Assumed-good’ reflection and transmission standards
Sometimes it is necessary to simply assume that a particular calibration standard is ‘good’. An example here might be an energy-absorbing termination (a ‘matched load’). This type of standard should produce no reflections and it is designed to absorb all of the incident energy. For the purposes of calibration, we must assume that this is the case.

The matched loads provided in a waveguide calibration kit are generally ‘good’ but not perfect. Here, ‘good’ typically means that the true reflection coefficient is likely to be in the region of 0.01 (at best) to around 0.03. For this reason, it is sometimes necessary to exercise caution when using matched loads as part of a calibration procedure (see the discussion below on calibration techniques).

Table 8.8 summarises the most common items found in a waveguide calibration kit. We must now consider the options available to the VNA metrologist for calibrating the VNA using these standards.

<table>
<thead>
<tr>
<th>Typical quantity</th>
<th>Standard</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 or 2</td>
<td>Flush short-circuits</td>
<td>Easy to implement in waveguide, high confidence in the reflection coefficient. Usually two are supplied as these are relatively easy to manufacture.</td>
</tr>
<tr>
<td>1 or 2</td>
<td>Shims (very short lengths of waveguide); where two shims are supplied, they will be different lengths.</td>
<td>For many calibration methods, the precise length needs to be known.</td>
</tr>
<tr>
<td>1 or 2</td>
<td>Offset short-circuits</td>
<td>Optional. Some calibration kits have a combined offset shim plus short-circuit, manufactured as a single device.</td>
</tr>
<tr>
<td>2</td>
<td>Matched loads (or sometimes, ‘low-reflecting loads’)</td>
<td>Must normally assume that the reflection coefficient is zero, unless using a calibration method that calls for a low-reflecting load (where the precise reflection value does not need to be known).</td>
</tr>
<tr>
<td>1</td>
<td>Verification lines – a length of waveguide, usually around 25 mm long. The length should be precisely known at standard lab temperature.</td>
<td>Not used in the calibration process, but often used to verify the success of a calibration.</td>
</tr>
</tbody>
</table>

Table 8.8: Typical standards likely to be found in a waveguide calibration kit. Other items may be found in specific calibration kits, but these are the main calibration standards used.
Fig 8.16: Common calibration standards: (a) flush short-circuit; (b) offset shim; (c) matched load.

8.3.3 One-port calibration methods

As discussed in the introduction to calibration principles above (8.3.1), for a one-port calibration, we require three known terminations (sometimes called a ‘three known loads’ calibration). The possible calibration strategies are therefore quite limited. The most common technique is the short/offset-short/load method. This is essentially a waveguide equivalent to the short/open/load (SOL) method used for coaxial calibrations at microwave frequencies. As observed earlier, open-ended waveguides are not commonly used at millimeter-wave frequencies and above. However, as with the lower-frequency short/open/load calibrations, the short and the offset-short in the waveguide ‘SOL’ calibration provide two strong reflections but with different phases.

It is important that, over the frequency range concerned, the length of the offset is such that the phase-shift produced is not the same as that obtained by a flush short-circuit (180°). Ordinarily, the length of the shims in the calibration kit are chosen such that either (a) this will not happen, or (b) by using two different shims, at least one will produce sufficient phase difference at every frequency in the band of interest.

Most VNAs calculate the actual phase-shift for each frequency point by first computing the delay at each frequency, using the cut-off frequency for the guide and the free-space delay. Therefore, it is important that the calibration shims are properly specified in terms of the minimum frequency, which must normally be equal to the cut-off frequency for the waveguide size (see equation 8.8), and the free-space delay.

A variation on this type of calibration calls for an additional measurement of an offset load. This approach is predicated on the fact that, very often, the electrical properties of the offset (i.e. a short length of waveguide with negligible attenuation and calculable delay) are more reliably known than

---

21 The acronym ‘SOL’ traditionally stands for ‘short/open/load’, although for calibration in waveguide, the open-circuit standard is substituted (without loss of accuracy) with an offset short-circuit. Usually, ‘SOL’ is still used to describe this type of calibration approach, although some manufacturers may use different acronyms, such as ‘SOSL’, or ‘SSL’.

22 The recommended minimum phase difference between the reflection coefficients of two standards is usually 20°.

23 This information is normally supplied with the calibration kit, and may already be installed in the VNA calibration kit definitions. It is prudent to check that the VNA has the correct values for minimum frequency and free-space delay for each standard (where appropriate) prior to starting a calibration. NB: Some modern VNAs now associate the minimum frequency with the test port waveguide size, rather than the individual calibration standards. The user can still specify a minimum frequency for the standards, but it is only used to inform the VNA of the preferred lower frequency at which the individual standards are to be used, rather than specify the cut-off frequency for the waveguide size.
the reflection properties of the load. Also, the load is usually not a perfect (i.e. reflection-less) load. In fact, for the offset-load calibration algorithm to work, it is actually preferable if the load is not absolutely perfect. To perform this type of calibration, a length of offset waveguide is required (in the same manner as the offset-short). The VNA measures the reflections from the load standards, with and without the offset, and then mathematically manipulates the results to determine a single, more accurate ‘measurement’ corresponding to an ideal reflection-less load. Again, it important that the phase shift produced by the offset line does not lead to phases which are coincident with the flush load.

An alternative to the ‘offset load’ calibration standard is a ‘sliding load’. These devices are commonly used at microwave frequencies where the load element is mounted on a coaxial, sliding assembly. By making multiple measurements at different positions, the reflection coefficient of a slightly-imperfect matched load traces a circle in the complex reflection coefficient plane. The centre of this circle corresponds to an ideal, reflection-less load. In metallic waveguides, the sliding load is formed by adding a micrometer adjustment to the energy absorbing element, such that its position (i.e. phase) may be continuously adjusted with respect to the test-port. Sliding loads are not easy to use (at any frequency) because a good calibration requires several, well-spaced measurements. In the hands of a skilled practitioner though, sliding loads can be used to achieve a good quality calibration.

It is actually possible to calibrate a one-port VNA without using a matched load. Providing two different offset lengths are available, a flush short plus two different offset-shorts may be used. However, for most purposes, a successful one-port calibration is achieved using the short/offset-short/load method and this is generally the recommended strategy.

Once calibrated using one of the methods described above, the VNA may then be used to make error-corrected, one-port measurements. The measured data will be mathematically corrected for all systematic magnitude and phase errors, and the magnitude and phase of the reflection measurement will be relative to the ‘reference plane’ established at the test-ports by the calibration standards.

8.3.4 Two-port calibration methods

For two-port calibrations, there are more options available. The most common strategies are members of either the short/offset-short/load/thru (SOLT)24 family of calibration methods, or the thru/reflect/load/thru (TRL) category [19].

SOLT
This calibration comprises a one-port calibration performed at each test-port, plus additional measurements for a ‘thru’ (through) connection between the two test-ports. The additional measurements are needed in order to allow the VNA to solve for the transmission tracking error terms.

Ordinarily, the ‘thru’ standard is implemented by simply connecting the waveguide test ports together. Effectively, this forms a zero-length path, with a predictable (i.e. zero) delay and perfect (i.e. lossless) transmission between the ports. Strictly speaking, the ‘thru’ does not need to be zero-length, but the transmission delay does need to be precisely known. Connecting the test-port reference planes together (yielding a zero-length path) makes this requirement easy to satisfy.

Some modern VNAs offer a development of the SOLT method which uses an ‘unknown thru’, where the precise characteristics of the ‘thru’ standard do not need to be known [20]. At first, this may appear to violate the basic principle of calibration methods discussed so far – the need to fully know the transmission/reflection properties of the calibration standards – but the method works by applying

---

24 At microwave frequencies (in coaxial line) this method is normally implemented as short/open/load/thru.
some elegant mathematics to the ‘raw’ measurement data in order to solve for the required error coefficients. In practice, some facts do need to be known about the ‘unknown thru’; it must be known that the ‘thru’ is reciprocal (i.e. $S_{21} = S_{12}$) and the user should be able to estimate the delay (or nominal phase shift). These are generally quite easy criteria to satisfy. When the unknown thru method is used, the calibration algorithm is often called short/offset-short/load/reciprocal thru (SOLR).

**TRL**
The second category of two-port calibrations is based on the thru/reflect/line (TRL) technique [19]. The TRL calibration algorithm operates in an entirely different manner from that of SOLT. Its main feature is that it cleverly avoids the need to know as much accurate information about the electrical properties of the calibration standards, when compared to SOLT.

TRL makes use of the fact that a two-port calibration is really a matter of establishing a calibration in an arbitrary wave-guiding structure (transmission medium), and the reflection/transmission behaviour of the DUT is relative to this medium. Thus, it is not actually necessary to know everything about the calibration standards, providing it is possible to use a section of the ‘reference’ wave-guiding structure as part of the calibration. In this way, TRL has been called a ‘self-calibration’ algorithm.

For the TRL method, the ‘thru’ is usually formed by a direct connection of the test-ports. The delay of the thru standard must be known accurately, but for a zero-path length this is easy to satisfy, in the same manner as for SOLT. The thru standard is also used to establish the test-port reference planes for the measurement, which are defined to be at the centre of the thru path.

The ‘reflect’ standard can be any reflecting device; all that needs to be known is the approximate phase of the reflection coefficient (to within $\pm 90^\circ$), and that it produces a reasonably strong reflection. The precise magnitude of the reflection coefficient does not need to be known. In practice, the most commonly used reflect standard is the flush short-circuit.

The ‘line’ standard is simply a short length of waveguide (for example, an offset shim). The delay/phase-shift produced by the line standard does not need to be known precisely (again, $\pm 90^\circ$ is sufficient). However, there is one very important proviso, which is that the phase-shift produced by the line standard must be different from the thru standard, and the phase difference must not be close to either $0^\circ$ or $180^\circ$. It is normally recommended that the phase shift lies between $20^\circ$ and $160^\circ$.

For lower frequencies, this is relatively easy to satisfy in metallic waveguide. A typical waveguide band covers an upper/lower frequency ratio of approximately 1.5:1. An offset line (shim) which is a quarter-wavelength (based on the guide wavelength) at the geometric mean of this band will have a phase shift of approximately $55^\circ$ at the lowest frequency of the band, and approximately $120^\circ$ at the upper frequency of the band.

As an example, consider the waveguide band covering 110 GHz to 170 GHz (WR-06/ WM-1651), for which the geometric mean frequency is 136.75 GHz. The cut-off wavelength is found directly from the broad-wall dimension of the guide, as $\lambda_c = 2a = 3.302$ mm, and the free-space wavelength at 136.75 GHz is 2.192 mm (using $c = 299 792 458$ m/s, and assuming $\varepsilon_r \approx 1$).

---

25 The calibration algorithm fails (numerically) if the phase shift is $0^\circ$ or $180^\circ$.

26 The frequencies which are deemed to be the upper and lower useable frequencies for a given waveguide size are rounded to convenient values. This means that the ratio of upper/lower frequency varies slightly among the common waveguide sizes.
Using equation 8.12 we find that the guide wavelength at 136.75 GHz is

\[ \lambda_g = \frac{\lambda}{\sqrt{1 - \left(\frac{\lambda}{\lambda_c}\right)^2}} = 2.932 \, \text{mm} \] (8.16)

Thus, a quarter-wavelength shim at this frequency will be \(2.932/4 = 0.733\) mm, producing a phase-shift of 90°. The guided wavelengths at 110 GHz and 170 GHz will be 4.822 mm and 2.085 mm respectively. The corresponding phase-shifts are therefore 54.7° and 126.6°.\(^{27}\)

If the required shim length for a quarter-wavelength at the geometric mean of 110 GHz to 170 GHz is only 0.733 mm, we may begin to perceive a possible problem with this approach at higher frequencies. For the WM-380 waveguide size (500 to 750 GHz) the equivalent shim would need to be approximately 160 μm in length. Such short ‘lengths’ mean that a \(\lambda/4\) offset waveguide is too thin to be considered practical. Effectively, such shims would be thin metallic foils with the waveguide flange details impressed, but they would be much too fragile to be considered reliable calibration standards.

The solution is to use longer ‘lines’ which are more physically robust. The penalty this incurs is that the bandwidth over which each shim satisfies the TRL phase-shift criteria, will be reduced, and therefore more than one line is needed to cover an entire waveguide band. Often, the lines are chosen to be \(3\lambda/4\) at the centre of two smaller frequencies ranges. Longer lines may be used if the frequency range is split into several suitable sub-bands \([21]\).

For example, calibrating over the entire frequency range for WR-03/WM-864, 220 GHz to 325 GHz, would require an impractical single line standard of around 362 μm. If this band is divided into two sub-bands (for the purpose of calibration) then lines of 1.290 mm and 1.000 mm will provide phases of nominally 270° in the centre of two sub-bands (220 GHz to 274 GHz, and 246 GHz to 325 GHz). Overlapping bands are chosen to ensure that all frequencies in the band may be calibrated with lines which remain well clear of the phase-shifts (0°, 180°) where the calibration algorithm does not work.

**LRL**

A variant of the TRL method is known as line/reflect/line (LRL) \([19]\). For LRL, the first ‘line’ takes the place of the TRL ‘thru’ standard, and therefore (just as with TRL) the phase-shift (delay) of this line must be accurately known. The second line need only be different from the first by the required 20° to 160°, and therefore a quarter-wavelength difference in the two lines is sufficient to meet this criterion.\(^{28}\) As for TRL, the delay/phase-shift of the second line only needs to be specified to within ±90°.

Using LRL means that only two shims of reasonable (i.e. practical) thickness need to be used, and these will be sufficient to calibrate across an entire waveguide band. However, care must be taken to specify the calibration standards. For LRL calibration, the VNA must compute the phase-shift for the

---

\(^{27}\) The value of the actual phase-shift for a given length of waveguide at a specific frequency may be found by several means. As discussed in 6.3.3., VNAs usually achieve this by first computing delay at each frequency, for which the cut-off frequency and the free-space delay are required, rather than the cut-off wavelength and the free-space wavelength. Whichever method is used, the precise value of free-space propagation is always required, including the effect of relative permittivity of the medium filling the guide (usually air). For lines which are multiple wavelengths long, these calculations often become numerically sensitive, and care must be taken to avoid a loss of accuracy through insufficient numerical precision.

\(^{28}\) It can be seen that TRL is essentially a ‘special case’ of LRL, where the first line (which must always be known) has a zero length.
first ‘line’ standard, using the cut-off frequency (minimum frequency) and the free-space delay. Thus, LRL suffers from the fact that the calibration accuracy depends on the accuracy with which the properties of the first line can be specified. The advantage of TRL is that a zero path length thru connection has a well-defined delay of exactly zero. For higher millimeter-wave frequencies, TRL with multiple lines may offer better results than LRL.

Some further comments on TRL/LRL calibration:
For TRL and LRL calibration, the measurement reference plane may be established using the first line standard or by using the reflect standard. For TRL, it is preferable to use the ‘thru’ (the first ‘line’ standard) since it is well known. If the first line standard is used in LRL, the measurement reference plane is established at the centre of the line, and therefore it relies on the accuracy with which this line is known.

The alternative is to use the reflect standard to set the measurement reference plane. In this case, the reflection phase of the reflect standard must be accurately known. Ordinarily, a flush short-circuit is used for the reflect standard, and therefore it is possible to specify the reflection phase with confidence. Often, the calibration kit definition simply requires the reflect standard to be defined as a short circuit – the reflection phase is then easily determined by the VNA. For LRL calibration, it is generally preferable to use the reflect standard to set the measurement reference plane.

In addition to specifying the minimum frequency for the offset shims (‘line’ standards), a maximum frequency may also be specified/required by some VNAs. Where this is the case, the maximum frequency is used to instruct the calibration algorithm to avoid using a particular standard at unsuitable frequencies, where the phase shift may be approaching 0° or 180°. (As before, the minimum frequency must correspond to the cut-off frequency for the waveguide size, and is used by the VNA to compute delay/phase-shift.)

TRL and LRL are two-port calibration techniques. As such, they require the VNA to be equipped with two transmission-reflection extender heads. However, once completed, TRL/LRL calibrations may be used to perform one-port measurements. Depending on the availability and quality of calibration standards for one-port calibration, a TRL/LRL calibration may well be a preferred option, even for one-port measurements.

Selecting a two-port calibration
The reader may well ask what factors should ultimately determine the choice of two-port calibration approach for a given measurement task. In practice, this usually comes down to (a) availability of suitable calibration artefacts, and (b) whether or not a particular strategy is practical.

In general, TRL (and its variants) will yield more accurate results. This is due to the fact that the standards do not have to be as precisely known compared with SSLT in order for the algorithm to successfully compute the error terms. Thus, at frequencies up to the region of 400 GHz, TRL (or LRL) is usually the best option. Beyond this, and at frequencies approaching 1 THz, it is often found that the line standards become difficult to manufacture. Multiple longer lines may be used (as mentioned above) but attenuation, and the fact that more steps are required in the calibration process, can mean that TRL/LRL methods are less attractive. Some manufacturers of THz calibration kits have returned to SSLT methods for calibration at THz frequencies.

For all waveguide-based calibrations, the VNA user may also be required to enter (or confirm) a value for the characteristic impedance, \( Z_0 \). In view of the difficulty of defining characteristic impedance for waveguide transmission lines, it is common practice to simply assign the value of \( Z_0 \) to 1. This means that any VNA display of actual impedance values (computed from the reflection measurements of the DUT) are effectively normalised values, relative to the transmission impedance of the waveguide. The
value entered for $Z_0$ does not affect the calibration in any way; it simply determines the scaling factor for the display of DUT impedance.

**Isolation (calibration for cross-talk errors)**

All two-port calibration techniques include an additional, optional stage of the calibration procedure called ‘Isolation’. The purpose of this step is to quantify the amount of signal leakage between the two ports (i.e. crosstalk). If some of the test signal can ‘leak’ between the two test-ports during a measurement, this amounts to an alternative signal path for transmission measurements. If the DUT has a very low transmission coefficient (low $S_{21}$ or $S_{12}$), this alternative signal path could become significant, or even dominate the measurement. However, in rectangular waveguide, signal leakage is usually very small, and for most VNAs, the signal leakage is less than $1 \times 10^{-5}$. Unless the DUT is expected to have a *very* low transmission characteristic, it is not necessary to conduct the isolation part of the calibration procedure. Indeed, many VNAs invite the user to omit the ‘Isolation’ part of a two-port calibration process (in which case this error source is assumed to be zero).

Table 8.9 summarises the most common methods of calibrating a VNA for millimeter-wave, submillimeter-wave and THz measurements.

<table>
<thead>
<tr>
<th>Method</th>
<th>Standards</th>
<th>Variations</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOL</td>
<td>Flush short, offset short, matched load</td>
<td>Might need more than one ‘offset’ to ensure phases are different at all frequencies.</td>
<td>Most common one-port calibration technique. Needs all standards to be known accurately. Offset short usually implemented by combining flush short with an offset shim.</td>
</tr>
<tr>
<td>SOL + offset load</td>
<td>Flush short, offset short, matched load, offset matched load</td>
<td>May also use ‘sliding load’ if available, but note that this is a different algorithm in the VNA.</td>
<td>Uses the offset load to improve accuracy by using additional mathematics to find the response of a ‘perfect’ load.</td>
</tr>
<tr>
<td>SOLT</td>
<td>Flush short, offset short, matched load, thru connection</td>
<td>As per SOL, SOLT may require more than one offset short.</td>
<td>Most common two-port calibration technique. Needs all standards (including thru) to be known accurately. May prove to be more practical at very high frequencies where offset line lengths are very short.</td>
</tr>
<tr>
<td>SOLR</td>
<td>Flush short, offset short, matched load, thru connection</td>
<td>Some older VNAs do not offer this method.</td>
<td>Recent addition to two-port calibration. Similar to SOLT except that the thru only needs to be reciprocal. Uses algorithm similar to TRL for the thru standard.</td>
</tr>
<tr>
<td>TRL</td>
<td>Flush short, thru connection (shim), offset line (shim)</td>
<td>Use single or multiple lines</td>
<td>Most accurate two-port calibration technique, providing lines lengths are practical.</td>
</tr>
<tr>
<td>LRL</td>
<td>Flush short, two offset lines (shims)</td>
<td>Uses two lines to calibrate over the full band.</td>
<td>Almost as accurate as TRL, but requires precise knowledge of the first line standard. Use the reflect standard to define the reference planes.</td>
</tr>
</tbody>
</table>

Table 8.9: Summary of methods used for waveguide calibration.

The process of removing systematic errors through calibration cannot be regarded as perfect. Some residual errors will always remain. Later in this chapter, we will look at the issue of *verification* (assessing the outcome of an attempt to calibrate the VNA) and quantifying the *uncertainty* associated with the measurements.
8.3.5  Good practice tips

Tip 1: A VNA should always be allowed to ‘warm up’ before use, especially when high quality measurements are needed. As a general rule, a VNA with millimeter-wave/submillimeter-wave or THz extender heads should be switched on for at least a few hours before use in order to thermally stabilise. Failing to do this will result in noticeable ‘drift’ in the measurement trace, particularly if the laboratory temperature is cool (<20°C). This is because the internal temperature in the VNA ‘box’ and the extender heads will increase considerably during the first couple of hours. Small changes in signal path dimensions due to thermal expansion can lead to noticeable changes, particularly in phase.

Tip 2: The VNA should ideally be set for the desired frequency range, number of frequency measurement points, averaging factor/IF bandwidth and test-port power levels (where it is possible to adjust) before commencing the calibration procedure. Some VNAs allow interpolation of calibration points, meaning that a narrower frequency range can be subsequently selected for measurement. The VNA will then attempt to calculate the error correction factors for any new frequency points created. Extrapolation of calibration range is not normally allowed, so if in doubt as to precisely what frequency range is needed, calibrate over the full bandwidth permitted by the waveguide size in use.

Tip 3: Calibration kits supplied with a VNA should be treated with great care and respect. The calibration standards determine the quality of the measurements, and as such they must be kept in pristine condition. Handle calibration standards with care. Small scratches on the flanges of calibration standards, especially on the surface of flush short circuits, can have a noticeable effect on the quality of a calibration. Usually, the calibration kit is a very expensive item. The standards should be reserved for their intended use in calibration, rather than for other purposes (for example, as convenient items for creating reflections or terminations when using multiport devices).

Tip 4: Take care making waveguide connections during the calibration process. For example, use all four screws on the waveguide flange (this can make a noticeable difference to the results). Remember that, as well as the condition of the standards, the quality of the calibration process will influence the quality of the ensuing measurements.

Tip 5: Always perform a ‘sanity check’ on the calibration before drawing conclusions from measurement results. It is not good practice to simply re-measure the same standards used in the calibration in order to verify the success of the calibration process. This is because, in many cases, the results will only confirm that the VNA can tell the user something that the user has just told the VNA! For example, if a poor quality matched termination is used in a one-port calibration, the same termination will look ‘perfect’ when re-measured. If possible, measure a different device, perhaps something which, after a few ‘good’ calibrations, the user knows what the expected result should be. For one-port measurements, a different flush short circuit is often a good choice (and readily available). Thus should show a total (100 %) reflection with a 180° phase angle. For two-port calibrations, a length of waveguide is the ideal choice. This should show almost zero reflections and almost perfect transmission. Be prepared to calibrate again if the magnitude of the transmission coefficient ($S_{21}$) is greater than 1 (> 0 dB)!

Tip 6: As mentioned previously, avoid excessive handling of certain items as this can lead to thermal changes. Metallic waveguide is a very good conductor of heat and so holding a waveguide device, such as a calibration load, can cause heat to be conducted inside the waveguide and onto the low-reflecting resistive element used to form the load. An increase in temperature can cause the reflection of the load element to change significantly and hence degrade the quality of the VNA calibration.
8.4 Errors and Uncertainties

8.4.1 Main sources of measurement error

Errors in VNA measurements are usually caused by imperfections in the system hardware. (When we talk about errors here, we are ignoring any blunders or incompetency on the part of the equipment operator.) These errors include the VNA and any accessories, as well as components such as the calibration standards and the devices under test. These imperfections can give rise to either random or systematic errors, or sometimes both.

We will discuss these sources of error under the following three headings:

- Waveguide transmission lines
- VNA hardware
- Calibration standards

Waveguide transmission lines – the waveguides that are used as part of the overall measurement setup will never be perfect. For example, for a waveguide aperture, the width, $a$, and height, $b$, of the aperture will not be exactly the correct size. In addition, the corners of the rectangular aperture will not be perfect right-angles. The lack of perfection in the waveguide flange alignment mechanisms (e.g. using dowel pins and holes) will cause some amount of misalignment, both linear misalignment in the $a$- and/or $b$-directions, and, angular misalignment. This misalignment will cause systematic errors, due to the alignment mechanisms (dowel pins and holes) not being exactly in the correct place, and random errors, due to the inevitable dimensional tolerances on these alignment mechanisms (e.g. the diameters of the dowel pins and holes). These waveguide dimensional errors will affect the performance of the VNA test ports (i.e. the waveguides on the extender heads attached to the VNA), the calibration standards, and, the devices under test.

VNA hardware – the sources of error in the VNA hardware are due to electrical noise, nonlinearity in the VNA’s detectors, and isolation/crosstalk between the VNA test ports. Electrical noise induces random errors into all VNA measurements. This can be further divided into trace noise and detector noise. Trace noise can be reduced by applying numerical averaging – i.e. by setting the VNA to take multiple measurements at the time that the VNA captures the data. Detector noise can be reduced by selecting a narrow intermediate frequency (IF) bandwidth, e.g. 100 Hz or less.

Nonlinearity in the VNA’s detectors will cause an error in the response of the VNA that varies as a function of measured signal level. For example, the error will increase as the amount of attenuation being measured increases. This nonlinearity error, $L$, is usually specified in terms of $L$ dB/dB. For example, if $L = 0.01$ dB/dB, then the nonlinearity error when measuring a 20 dB attenuator will be $0.01 \times 20 = 0.2$ dB.

Isolation/crosstalk between the VNA’s test ports means that not all the test signal that is supplied by the VNA actually passes through the device under test. Instead, some of the signal is picked up directly by the VNA’s receiver detectors without passing through the device under test. This can cause problems, especially when the VNA is measuring low signal levels (e.g. for devices with high attenuation).

Calibration standards – The assumptions made about the properties of calibration standards will never be exactly true. For example, it is common to assume that the ‘matched’ load standard really is matched (i.e. produces no reflected signal). In practice, there will always be some signal reflected by the ‘matched’ load and so the assumption that this reflected signal is actually zero will induce an error
into the calibration process. This calibration error will then induce an error in all measurements made with respect to this calibration (using this `matched’ load standard).

Similarly, if an offset device is used during calibration (e.g. an offset short-circuit standard) then the amount of offset (i.e. the length of line inherent in the standard) will not be known perfectly. The error in the knowledge about this length will induce an error into the calibration process. As before, this calibration error will then induce an error in all measurements made with respect to this calibration (using this offset short-circuit standard).

8.4.2 Typical sizes of errors

Much work has been done, over many years, to investigate waveguide transmission line errors due to waveguide dimensional tolerances. These days, a convenient method of evaluating such errors is by using electromagnetic simulation software. A report showing the results produced by one such simulator is available at: [http://www.gb.nrao.edu/electronics/edtn/edtn215.pdf](http://www.gb.nrao.edu/electronics/edtn/edtn215.pdf) [22]. This report shows calculations of mismatch due to tolerances on aperture height and width dimensions, and, aperture corner radii. The report also shows calculations of mismatch due to both linear and angular flange misalignments.

The report by Kerr [22] shows that, for good quality waveguide, the most significant sources of mismatch are likely to be due to tolerances on the height and width dimensions of the waveguide aperture as well as linear flange misalignments (in both the a- and b-directions). Approximate analytical formulas for these effects have been given in [23]. For example, the mismatch (in terms of the linear magnitude of the reflection coefficient, $|\Gamma|$) due to a height error, $\delta b$, in a waveguide aperture of height, $b$, can be approximated by:

$$|\Gamma| = \frac{|\delta b|}{2b}$$

(8.17)

Similarly, the linear magnitude of the reflection coefficient, $|\Gamma|$, due to a width error, $\delta a$, in a waveguide aperture of width, $a$, can be approximated by:

$$|\Gamma| = \frac{1}{8} \left( \frac{\lambda_g}{a} \right)^2 \frac{|\delta a|}{a}$$

(8.18)

where $\lambda_g$ is the guide wavelength.

We can therefore use equations 8.17 and 8.18 to estimate the mismatch (in terms of worst-case reflection coefficient) caused by different waveguide aperture height and width errors. Table 8.10 shows this information (in dB) for aperture height errors, $|\delta b|$, of 1 μm, 2 μm, 5 μm and 10 μm. Similarly, Table 8.11 shows similar information (in dB) for aperture width errors, $|\delta a|$, of the same size.

| Waveguide Name | Minimum frequency (GHz) | Maximum frequency (GHz) | $|\delta b| = 1 \mu m$ | $|\delta b| = 2 \mu m$ | $|\delta b| = 5 \mu m$ | $|\delta b| = 10 \mu m$ |
|----------------|-------------------------|-------------------------|-----------------|-----------------|-----------------|-----------------|
| WM-570        | 330                     | 500                     | -54             | -48             | -41             | -35             |
| WM-380        | 500                     | 750                     | -50             | -46             | -38             | -32             |
| WM-250        | 750                     | 1100                    | -48             | -42             | -34             | -28             |

Table 8.6: Estimated worst-case reflection coefficient (from [23]) due to errors in waveguide aperture height of 1 μm, 2 μm, 5 μm and 10 μm.
<table>
<thead>
<tr>
<th>Waveguide Name</th>
<th>Minimum frequency (GHz)</th>
<th>Maximum frequency (GHz)</th>
<th>Worst-case reflection coefficient (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM-570</td>
<td>330</td>
<td>500</td>
<td>-54  -50  -42  -36</td>
</tr>
<tr>
<td>WM-380</td>
<td>500</td>
<td>750</td>
<td>-54  -48  -39  -33</td>
</tr>
<tr>
<td>WM-250</td>
<td>750</td>
<td>1100</td>
<td>-48  -43  -35  -29</td>
</tr>
</tbody>
</table>

Table 8.7: Estimated worst-case reflection coefficient (from [23]) due to errors in waveguide aperture width of 1 μm, 2 μm, 5 μm and 10 μm.

The errors in the VNA hardware – noise, nonlinearity, isolation/crosstalk – vary significantly depending on the type of VNA and the VNA’s operating conditions (including frequency range, etc). Errors due to noise can be reduced by using numerical averaging and a narrow IF bandwidth. Effects due to noise can often be taken into account at the same time as other random effects are evaluated (e.g. the errors due to flange connection repeatability – see subsection 8.4.3, below).

Nonlinearity errors can usually be evaluated by measuring a series of ‘known’ values of attenuation. At lower frequencies, a calibrated step attenuator is often used to provide a series of increasing values of attenuation. Measurement of these attenuation ‘steps’ helps to establish if there is an error that increases proportionally to the level of attenuation being measured. At the present time, there are no internationally agreed methods for evaluating nonlinearity errors in VNAs operating at submillimeter-wave frequencies. However, work is on-going in this area at National Measurement Institutes (such as NIST in the USA, NPL in the UK, and PTB in Germany) and so these institutes should be consulted for the most up-to-date information about this topic.

Isolation/crosstalk errors can be evaluated by measuring the transmission (i.e. $|S_{21}|$ and $|S_{12}|$) when both VNA ports are terminated with low reflecting loads. In an ideal world, no signal should be passing from port 1 to port 2 (as no connection has been made between ports 1 and 2 and so no signal path has been made available for transmission between the ports). However, in practice, some signal coming from the VNA’s source circuitry can get picked up by the VNA’s receiver circuitry due to leakage paths between source and receiver. Figure 8.17 shows some typical results of the measured transmission for a VNA operating from 750 GHz to 1100 GHz and terminated with two low reflecting loads. On this occasion, the VNA’s IF bandwidth was set to 30 Hz. This helps reduce any effects due to noise on the observed isolation/crosstalk.

![Figure 8.17: Typical forward ($|S_{21}|$) and reverse ($|S_{12}|$) isolation/crosstalk for a VNA operating in the 750 GHz to 1100 GHz waveguide band.](image-url)
The amount of isolation/crosstalk in a VNA system is often dependent on the operating frequency – there is usually more crosstalk present at very high frequencies compared with lower frequency ranges. For example, Figure 8.18 shows the measured crosstalk/isolation for a VNA operating from 140 GHz to 220 GHz, where the isolation/crosstalk is seen to be typically within the range -70 dB to -90 dB. This can be compared with the isolation/crosstalk shown in figure 8.17, where the VNA is operating from 750 GHz to 110 GHz and the measured crosstalk/isolation is seen to be typically from -40 dB to -60 dB.

Figure 8.18: Typical forward (|S_{21}|) and reverse (|S_{12}|) isolation/crosstalk for a VNA operating in the 140 GHz to 220 GHz waveguide band.

8.4.3 Connection repeatability

As mentioned in subsection 8.2.4 (Flanges/Interfaces), waveguide flanges can be a significant source of errors in VNA measurements. These errors can be both systematic in nature (e.g. due to a consistent misalignment between the two connecting waveguides) or random (e.g. due to dimensional tolerances on the alignment mechanisms and other random effects). The random errors give rise to measurements that do not repeat exactly each time the waveguide flange connection is disconnected and re-connected.

Random errors due to connection repeatability can be assessed by repeatedly disconnecting and reconnecting a device a number of times, and analysing the results in terms of the variability in the results, i.e. using statistical techniques. Some typical repeatability assessments using two VNAs – one operating in the 140 GHz to 220 GHz band and another operating in the 750 GHz to 1100 GHz band – are shown in Tables 8.12 and 8.13, respectively. These Tables give (at the minimum, middle and maximum frequencies for each waveguide band) the experimental standard deviation in the magnitude of the reflection coefficient, s(|\Gamma|), determined from repeated disconnects/reconnects of two devices: (i) a near-matched load, and (ii) an offset short-circuit. For the assessment shown here, 12 disconnect/reconnects of each device were used to calculate s(|\Gamma|) according to the following formula

\[
s(|\Gamma|) = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} |\Gamma_k - \Gamma|^2}
\]  

(8.19)
where \( n \) is the number of repeat measurements (in our case, \( n = 12 \)), \( \Gamma_k \) \((k = 1 \text{ to } 12)\) are the repeat measurements of the complex-valued reflection coefficient and \( \bar{\Gamma} \) is the complex-valued mean reflection coefficient:

\[
\bar{\Gamma} = \frac{1}{n} \sum_{k=1}^{n} \Gamma_k + j \frac{1}{n} \sum_{k=1}^{n} \Gamma_k
\]

(8.20)

where \( \Gamma_R \) and \( \Gamma_I \) are the real and imaginary components of the reflection coefficient, respectively.

<table>
<thead>
<tr>
<th>Frequency (GHz)</th>
<th>Near-matched load</th>
<th>Offset short-circuit</th>
</tr>
</thead>
<tbody>
<tr>
<td>140</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>180</td>
<td>0.002</td>
<td>0.007</td>
</tr>
<tr>
<td>220</td>
<td>0.002</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 8.12: Repeatability assessments, in terms of \( s(|\Gamma|) \), for the WR-05 waveguide band.

<table>
<thead>
<tr>
<th>Frequency (GHz)</th>
<th>Near-matched load</th>
<th>Offset short-circuit</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>0.012</td>
<td>0.136</td>
</tr>
<tr>
<td>900</td>
<td>0.009</td>
<td>0.093</td>
</tr>
<tr>
<td>1100</td>
<td>0.011</td>
<td>0.078</td>
</tr>
</tbody>
</table>

Table 8.13: Repeatability assessments, in terms of \( s(|\Gamma|) \), for the WR-01 waveguide band.

During the repeatability assessments shown here, the flanges on the devices under test were maintained with the same orientation with respect to the VNA test port flanges (i.e. the devices were deliberately not inverted as this would introduce an additional source of variability in the measurements). This repeatability assessment therefore only responds to random effects such as the tolerances on the diameters of the dowel pins and holes used to align the waveguides. If the devices are inverted during the repeatability assessment then the effects to due incorrect position of the dowel pins will also have an impact in the observed repeatability. In general, including flange inversion in the repeatability assessment will increase the amount of variability (and hence the observed standard deviations) in the measurements. This will be particularly true at the very high frequencies (e.g. towards 1 THz, and above).

Tables 8.12 and 8.13 show that, for the same type of device (i.e. either matched load or short-circuit) the \( s(|\Gamma|) \) values at the lower frequencies (140 GHz to 220 GHz) are smaller than at the higher frequencies (750 GHz to 110 GHz). This decrease in repeatability at the higher frequencies is most likely due to waveguide aperture at these frequencies being much smaller than the waveguide aperture for the lower frequencies, while the dimensional tolerance errors affecting the flange alignment mechanism for both waveguide sizes are effectively the same. A similar flange dimensional misalignment error will therefore have a much greater impact on the smaller waveguide aperture compared to the larger waveguide aperture.

Tables 8.12 and 8.13 also show that, for a given waveguide size (either WR-05, 140 GHz to 220 GHz, or WM-250, 750 GHz to 1100 GHz), the \( s(|\Gamma|) \) values are smaller for the near-matched load compared with the offset short-circuit. This shows that the measurement repeatability is also related to the size (i.e. magnitude) of the signal being measured – for the near-matched load, the amount of reflected signal is very low (i.e. close to zero); for the offset short-circuit, the amount of reflected signal is very high (close to 100% reflected signal).
Finally, it is worth noting that the standard deviation calculation shown in equation 8.19 does not show separately the variation in the real and imaginary components of the complex-valued reflection coefficient. Rather, the equation ‘merges’ the variability of the two components into one. In order to examine separately the variation in the real and imaginary components, the calculation should be performed separately on these two components. Advice on performing such a calculation has been given in [24].

8.4.4 System verification

After calibrating the VNA system, it is important to perform some kind of verification on the overall system performance to check that assumptions made during the calibration process (e.g. about the system and/or the calibration standards) were correct. For example, that the amount of reflected signal provided by the near-matched load was sufficiently small.

Until recently, it has been very difficult to independently verify the performance of a VNA at these high millimeter- and submillimeter-wave frequencies. This is because devices providing known values (against which the VNA can be checked) have not been readily available. At lower frequencies, manufacturers provide verification kits for the VNA but at the present time, such kits are not readily available at the very high frequencies.

However, some progress has recently been made to address these problems. A new form of ‘standard’ has been developed that comprises a straight section of waveguide that is orientated during connection such that the waveguide aperture is at right-angles to the waveguide apertures on the VNA test ports. This has been described elsewhere in [25, 26]. This ‘cross-connected’ waveguide (or ‘cross-guide’, for short) forms a section of waveguide that is effectively below cut-off and so its loss can be predicted from electromagnetic theory, e.g. using 3-D electromagnetic simulation software.

A cross-guide device can be connected to the VNA after calibration to verify the VNA’s performance, by comparing the VNA results with values predicted by simulation software, or, with measured values supplied by another laboratory (such as a standards laboratory). The level of agreement is established by comparing the difference between the two sets of results with the expected uncertainty in the results. For example, figure 8.19 shows the measured and modelled values of transmission coefficient for two different lengths of cross-guide in WR-05 (140 GHz to 220 GHz): 0.54 mm and 1.47 mm. Figure 8.20 illustrates the concept of a cross-connected waveguide and figure 8.21 illustrates the connection strategy for these cross-guide lines.
Figure 8.19: Measured (red curve) and modelled (blue curve) transmission coefficient values for a 0.54 mm length of WR-05 cross-guide and a 1.47 mm length of WR-05 cross-guide.

Figure 8.20: Illustration showing the concept of a cross-connected waveguide or ‘cross-guide’.

Figure 8.21: (a) a conventional waveguide test port; (b) a waveguide line orientated as a cross-connected waveguide (so that its aperture is rotated 90° with respect to the test port aperture).
8.4.5 Good practice tips

Tip 1: Devices used for verification purposes should be connected with the same degree of care and attention as the calibration standards and devices under test. A poor connection of a verification device could lead to the conclusion that the VNA is not operating correctly when actually it is the verification device that is not operating as it should (due to the poor connection).

Tip 2: Always ensure waveguide flange faces are clean. This can be established by visual inspection – a hand-held magnifier can be used to help with the inspection of the flange faces. Any dirt can be removed using a cotton swab dipped in a suitable cleaning fluid (i.e. Isopropanol (IPA)).

Tip 3: Keep documented records of the performance of the VNA - for example, periodic measurements (or plots) of the VNA’s isolation/crosstalk. These records can be used on subsequent occasions to show that the VNA is performing as it did previously.

Tip 4: If possible, try to obtain some dimensional data on the critical waveguides used with the system. For example, the aperture dimensions (height and width) of the calibration standards and the VNA extender head test ports (i.e. the calibration reference planes). Check the measured dimensions against values given in the standards [16].

8.5 Concluding remarks

This chapter has given an overview of the Vector Network Analyzer (VNA) as used at millimeter and submillimeter wavelengths. Emphasis has been given on using these VNAs for making measurements in rectangular metallic waveguide. This is because the extender heads that are used with these VNAs at these frequencies use this type of waveguide to provide the test ports. A description of the operation of the VNA and its hardware has been given along with an overview of the properties of metallic waveguides used at these frequencies.

Information has also been given concerning the use of standards and calibration procedures that are required in order for the VNA to make meaningful measurements. An uncalibrated VNA is rarely of any use in the majority of measurement applications; whereas a calibrated VNA that has used appropriate standards and calibration routines is capable of making very accurate and very reliable measurements. Some information has been given on understanding the errors and uncertainties associated with VNA measurements. Ultimately, in the fullness of time, VNA measurements made at these frequencies should be made traceable to the international system of units (SI). Although some work has now begun in this area [27] there is still much more that needs to be done before VNA measurements can be considered ‘traceable’ at all these frequencies.

Finally, it is worth noting that measurements using VNAs can also be made in other transmission media (e.g. in planar circuit, on-wafer, environments) at millimeter and submillimeter wavelengths. This is achieved by attaching adaptors (e.g. waveguide to co-planar waveguide on-wafer probes) to the waveguide ports of the VNA extender heads. However, this topic is outside the scope of the material covered by this chapter. At the present time, measurements using on-wafer probes, at these frequencies, is a subject that is still in its infancy. More work needs to be done by researchers and manufacturers before this technology can be taken up and implemented by the on-wafer user community.
References


