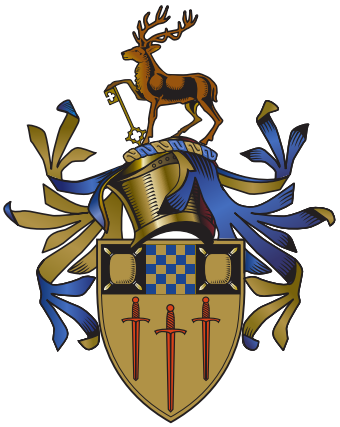
MEASUREMENT UNCERTAINTY IN NON-LINEAR BEHAVIOURAL MODELS OF MICROWAVE AND MILLIMETRE-WAVE AMPLIFIERS

Laurence Stant



A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Advanced Technology Institute and Department of Electronic Engineering
Faculty of Engineering and Physical Sciences
University of Surrey

September 2017

Declaration of Authorship

I confirm that the submitted work is my own work and that I have clearly identified and fully acknowledged all material that is entitled to be attributed to others (whether published or unpublished) using the referencing system set out in the programme handbook. I agree that the University may submit my work to means of checking this, such as the plagiarism detection service Turnitin UK. I confirm that I understand that assessed work that has been shown to have been plagiarised will be penalised.

The authors confirm that data underlying the findings are available without restriction. Details of the data and how to request access are available from the University of Surrey publications repository at http://epubs.surrey.ac.uk/blahTBD

Laurence Stant (Author)	Date

"What error drives our eyes and ears amiss? Until I know this sure uncertainty I'll entertain the offered fallacy."

William Shakespeare, The Comedy of Errors

"That's right!" shouted Vroomfondel, "we demand rigidly defined areas of doubt and uncertainty!"

Douglas Adams, The Hitchikers Guide to the Galaxy

Abstract

Abstract goes here

Research Outcomes

Publications

[1] M. Salter, L. Stant, K. Buisman, and T. Nielsen, "An inter-laboratory comparison of NVNA measurements," in *2011 Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits*, Brive, Jul. 2018.

[2] L. Stant, M. Salter, N. Ridler, D. Williams, and P. Aaen, "Propagating measurement uncertainty to microwave amplifier nonlinear behavioural models," *IEEE Trans. Microw. Theory Techn.*, Nov. 2018.

[3] L. Stant, P. Aaen, and N. Ridler, "Evaluating residual errors in waveguide VNAs from microwave to submillimetre-wave frequencies," *IET Microwaves Antennas Propag.*, vol. 11, no. 3, pp. 324–329, Feb. 2017. DOI: 10.1049/iet-map.2016.0455.

[4] ——, "Comparing methods for evaluating measurement uncertainty given in the JCGM 'evaluation of measurement data' documents," *Measurement*, vol. 94, pp. 847–851, Dec. 2016. DOI: 10.1016/j.measurement.2016.08.015.

[5] ——, "Evaluating residual errors in waveguide network analysers from microwave to submillimetre-wave frequencies," in *IET Colloquium on Millimetre-Wave and Terahertz Engineering & Technology 2016*, Institution of Engineering and Technology (IET), 2016. DOI: 10.1049/ic.2016.0016.

Acknowledgements

I would like to thank my supervisory team: Peter Aaen, Nick Ridler and Marian Florescu. Peter and Nick gave me many opportunities throughout the project and were especially helpful and encouraging when problems arose. Marian took over as my primary supervisor while writing this dissertation and offered great advice and support to help me complete my work.

Martin Salter at the National Physical Laboratory (NPL), Teddington, who learned about nonlinear behavioural models and their extraction alongside me, always provided fruitful discussions and these helped me a great deal in my understanding of the science and mathematics. Martin also assisted with or performed some of the measurements in our publications.

I would also like to thank Daniel Stokes and other scientists in the electromagnetic measurements labs at NPL, who provided training and access to use their equipment, including the National primary standards for impedance in a variety of coaxial and waveguide media. Their welcoming and friendly nature also made every trip to NPL interesting, even when the schedule consisted entirely of repeated measurements!

I must extend huge thanks to Dylan Williams and his group at the National Institute for Standards and Technology (NIST) in Boulder, CO, USA. When I made contact with Dylan in regards to extending his existing software uncertainty framework he was immediately helpful and encouraging. He extended his upcoming UK trip to European Microwave Week and stayed at the university for an additional week to work with me expanding his software. The following year he invited me to visit NIST for three weeks to spend more time integrating behavioural model extraction into his software, and to characterise phase references used in my experiments. Both Dylan and his colleagues were excellent hosts and it was really exciting to work in an office with so much knowledge and experience. Special thanks also to Gustavo Avolio who performed the characterisation of our phase references, and also reviewed my final publication.

Away from my desk, my mother has provided me with a lot of support during my studies and put up with my various complaints. She has never stopped believing that I could finish my project and must feel quite vindicated now!

I am enormously grateful to my colleague and friend throughout my entire university career, Sean Gillespie. Together we were Peter's first PhD students at the University of Surrey, and experienced all the ups and downs of starting a new lab. We saw the university change a lot over the previous nine years, and I'm sure we will both look back in future with very fond memories of it.

Finally, last but not least, I would like to thank my fellow students and staff at the Advanced Technology Institute at the University of Surrey: Haris, Gemma, Jonas, Morgan, Steve, Matei, Tony, John, Grace, Julian, Jose, Kostis, Scott, James, Bob - the list goes on. Together they made my time as a postgraduate thoroughly enjoyable and I will miss them all!

Contents

List of Figures

List of Tables

1 Introduction

1.1 Wireless Communications

1.2 Amplifier Measurement and Modelling

1.3 The Role of Uncertainty in Measurement

1.4 Thesis Structure

Testing, testing[1], [2].

Bibliography

[1] L. Stant, P. Aaen, and N. Ridler, "Evaluating residual errors in waveguide network analysers from microwave to submillimetre-wave frequencies," in *IET Colloquium on Millimetre-Wave and Terahertz Engineering & Technology 2016*, Institution of Engineering and Technology (IET), 2016. DOI: 10.1049/ic.2016.0016.

[2] ——, "Comparing methods for evaluating measurement uncertainty given in the JCGM 'evaluation of measurement data' documents," *Measurement*, vol. 94, pp. 847–851, Dec. 2016. DOI: 10.1016/j.measurement.2016.08.015.

2 Radio Frequency and Microwave Measurements

2.1 Introduction

To characterise nonlinear behavioural models, the radio frequency (RF) response of a device to electromagnetic wave stimuli must be measured. When compared with DC (and low frequency) measurements, RF and microwave measurements present significant additional challenges. For DC systems, it is desirable to propagate voltages through a circuit with minimal loss in amplitude. To achieve this effectively, components are typically designed with high input impedance and low output impedance. With RF systems, circuit components and interconnects can be of the order of a quarter-wavelength in length, and therefore signals must be treated as electromagnetic waves to account for different behaviour at these frequencies. When a travelling wave encounters a discontinuity in impedance, such as a cable connector or on-wafer structure, some of the power in the wave is reflected. The amount of reflected power is proportional to the size of the impedance mismatch between each side of the discontinuity. Hence, for RF systems, the transmission of power is the focus of the circuit designer. The measurement of power flowing through a transmission line is complicated by three key factors. Firstly, because the waves are travelling, the instantaneous voltage at any point on the transmission line will vary between the peak-to-peak values of the wave. Secondly, there are waves travelling in both directions along the transmission line which must be measured separately. Finally, the power of the wave is a complex quantity which consists of both magnitude and phase. To perform these measurements, a specialist instrument called a vector network analyser (VNA) can be used. In this chapter, the concepts and measurements associated with this instrument are introduced, which will be used later in the thesis to understand the uncertainty contributions from measurements to nonlinear

behavioural models.

2.2 Electromagnetic Wave Parameters

2.2.1 Wave Definitions

To describe the power of electromagnetic waves propagated through a transmission line, several definitions are in use in industry and academia for either accuracy or convenience. To avoid confusion in this document, these will now be defined. Information presented in this section has been obtained from [1-5].

2.2.1.1 Travelling Waves

Travelling waves represent a solution to Maxwell's equations along a transmission line. They are physical and measurable via slotted line experiments or thru-reflect-line calibrations [6] (see Chapter X). Travelling waves are defined by the total transverse electric and magnetic fields $\boldsymbol{E}_{\mathrm{t}}$ and $\boldsymbol{H}_{\mathrm{t}}$ of a single propagating mode at each frequency:

$$\boldsymbol{E}_{\mathrm{t}} = c^{+}e^{-\gamma z}\boldsymbol{e}_{\mathrm{t}} + c^{-}e^{+\gamma z}\boldsymbol{e}_{\mathrm{t}}, \quad \boldsymbol{H}_{\mathrm{t}} = c^{+}e^{-\gamma z}\boldsymbol{h}_{\mathrm{t}} - c^{-}e^{+\gamma z}\boldsymbol{h}_{\mathrm{t}} \tag{2.1}$$

where, following the notation of [5], $\boldsymbol{e}_{\mathrm{t}}$ and $\boldsymbol{h}_{\mathrm{t}}$ are the un-normalized electric and magnetic fields of the modal solution of Maxwell's equations in transmission line, $\gamma = a + ib$ is the complex propagation constant of the mode, z is the direction of propagation, and c^{+} and c^{-} are complex quantities representing the un-normalized forward and backward amplitude of the mode, respectively.

2.2.1.2 Equivalent-Circuit Voltage and Current

To represent travelling waves as equivalent low frequency circuit parameters such as voltage and current, a normalisation is chosen to derive a characteristic impedance for the transmission line. This normalisation takes the form

$$\boldsymbol{E}_{\mathrm{t}}(z) = \frac{v(z)}{v_0}\boldsymbol{e}_{\mathrm{t}}, \quad \boldsymbol{H}_{\mathrm{t}}(z) = \frac{i(z)}{i_0}\boldsymbol{h}_{\mathrm{t}}, \tag{2.2}$$

where v_0 and i_0 are normalisation constants that allow v and i to take units of root-mean-square voltage and current, respectively [5].

2.2.1.3 Pseudowaves

Equivalent voltages and currents cannot be used in lossy transmission lines where the electric and magnetic fields are out of phase. To account for this and provide a solution which can be used with conventional circuit design methodologies (e.g. Smith chart techniques [7]) and simulators, pseudowaves can be used. This representation is defined with a reference impedance, Z_{ref}, which can be chosen by the user, but is typically 50-Ω in conventional measurements. The forward and backward pseudowaves a and b can be written as:

$$a(Z_{\text{ref}}) = \left[\frac{|v_0|}{v_0}\frac{\sqrt{\Re(Z_{\text{ref}})}}{2|Z_{\text{ref}}|}\right](v + iZ_{\text{ref}}), \quad b(Z_{\text{ref}}) = \left[\frac{|v_0|}{v_0}\frac{\sqrt{\Re(Z_{\text{ref}})}}{2|Z_{\text{ref}}|}\right](v - iZ_{\text{ref}}) \tag{2.3}$$

2.2.1.4 Power Waves

Finally, power waves are defined so that the relationship $P = |a|^2 - |b|^2$ is true for any reference impedance, where P is the power transmitted through the transmission line and a and b are the forward and backward power waves, respectively. They are defined as:

$$a(Z_{\text{ref}}) = \frac{v + iZ}{2\sqrt{\Re(Z_{\text{ref}})}}, \quad b(Z_{\text{ref}}) = \frac{v - iZ}{2\sqrt{\Re(Z_{\text{ref}})}}. \tag{2.4}$$

Data taken from Keysight instruments used later in this work is presented in power wave format, with units of square-root Watts. To convert these values into decibels referenced to 1 milliwatt, the following formula is used:

$$P(\text{dBm}) = 10\log_{10}(P(\sqrt{W})^2) + 30 \tag{2.5}$$

2.2.2 Derived Metrics and Figures of Merit

The behaviour of a linear microwave device can be completely defined by the complex ratio of electromagnetic waves which are scattered at each port to those which are incident at each port. The combination of these ratios constitutes the scattering parameters (s-parameters) of a microwave device and are used extensively in the design and measurement of microwave systems[ref]. The formal definitions of the s-parameters for a two-port device are

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0}, \quad S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}, \quad S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0}, \quad S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0}, \tag{2.6}$$

where both a and b can be expressed in either pseudowave or power wave representation. The term “scattered” can be interchanged with “transmitted” and “reflected” depending on if the scattered wave is output on a different port, or the same port, to the incident wave, respectively.

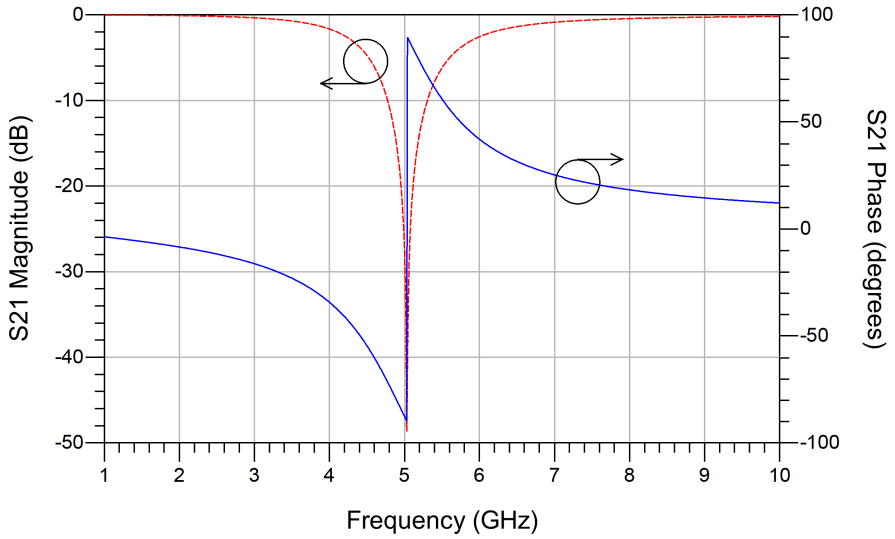


Figure 2.1: The frequency dependence of the magnitude (red dotted trace) and phase (blue solid trace) of S_{21} for a bandstop filter.

A signal flow diagram is provided in Fig. 1 showing the relationship between equivalent-circuit voltage and current, pseudowaves/power waves and s-parameters for a two-port device. The s-parameters of all microwave devices will exhibit some degree of frequency dependence. This effect originates from physical processes occurring in the device and can either be a benefit or hinderance to a design. Most passive components (including cables) will have a usable bandwidth which is an unwanted limitation, whereas microwave filters are a ubiquitous component where the same fixed bandwidth is the main purpose of the device. To capture this frequency dependence, s-parameters are measured across a frequency range and stored in a table, usually in Touchstone format (see Fig.2). An example of the frequency dependence of a filter is shown in Fig. 3. For a device operating in the linear regime, if multiple stimuli at different frequencies are incident on the device, they will not interact with each other. The scattered waves will have the same frequency components as if the stimulus at each frequency was applied separately. This is called the frequency superposition principle and does not apply to nonlinear operating regimes, which will be discussed later in this chapter.

Scattering parameters are often expressed in matrix form, where the column index is the scattered port, and the row index is the incident port. For a two-port device, the s-parameter

```
!Hewlett Packard, E8753E, US38432892
!Date: Saturday, February 03, 2018 19:29:08
!Correction: S11(Off) S21(Off) S12(Off) S22(Off)
!S2P File: Measurements: S11, S21, S12, S22:
# Hz S RI R 50
3.00000000000000E+4 9.34997558593750E-1 5.69458007812500E-2 -1.73002481460571E-5 -1.6689300537109
6.00000000000000E+9 1.53465270996094E-2 -6.19812011718750E-2 1.51917338371277E-5 5.52944839000702
```

Figure 2.2: An short example Touchstone file showing a two-port measurement at two frequencies. The rows continue to the right of the figure.

matrix would be

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \tag{2.7}$$

The most interesting characteristic of a two-port microwave device is often the effect which it has on a transmitted wave in the forward direction (S_{21}). If the device increases the magnitude of the incident signal this metric is called gain, otherwise it is called insertion loss. Typically gain is associated with active devices (those which are powered from an external source separate to the incident microwave signals) such as amplifiers, and insertion loss is associated with passive devices (those with no external power source) such as attenuators, splitters and mixers. The power gain (operating gain) and insertion loss relating to S_{21} can be calculated using

$$\text{Power Gain} = 10 \log_{10} |S_{21}|^2 \text{ dB}, \tag{2.8}$$

and

$$\text{Insertion Loss} = -10 \log_{10} |S_{21}|^2 \text{ dB}, \tag{2.9}$$

respectively.

Optimal transmission in microwave systems requires impedance matching between components, and it is inevitable that this matching will not be perfect and so some power will be reflected in a two-port device. Therefore, the match of a device is another important measurement, which is dependent on the voltage reflection coefficient (Γ) of the device and can be related to the impedance of a source and load by

$$\Gamma_{xx} = S_{xx} = \frac{Z_\text{L} - Z_\text{S}}{Z_\text{L} + Z_\text{S}}, \tag{2.10}$$

where x is a port index. A more thorough definition of voltage reflection coefficient for a two-port device includes any effect from the impedance seen at the other port, and for the case of input match is calculated as

$$\Gamma_{11} = S_{11} + \frac{S_{12} S_{21} \Gamma_\text{L}}{1 - S_{22} \Gamma_\text{L}}, \tag{2.11}$$

where Γ_L is the voltage reflection coefficient of the load connected to the device. For amplifiers, the amount of isolation (reduction of S_{12}) is an important characteristic of the device, whereby a fully isolated amplifier ($S_{12} = 0$) is said to be unilateral and equations 2.10 and 2.11 are equivalent.

For active devices, such as amplifiers, it can also be useful to consider the power reflected at the input when calculating the power gain of the device. The transducer gain of a device accounts for this potential loss of power at the input and provides a more portable metric which is not dependent on the impedance of the measurement setup. It is defined as

$$G_T = \frac{1 - |\Gamma_S|^2}{1 - |\Gamma_{in}\Gamma_S|^2} |S_{21}|^2 \frac{1 - |\Gamma_L^2|^2}{1 - |S_{22}\Gamma_L|^2}, \tag{2.12}$$

where Γ_{in} is the input match of the device.

For all devices operating in the linear regime, any reflected or transmitted wave will have a frequency equivalent to that incident to the device. In addition, the stimulus power that was used to measure the s-parameters is not important as the ratio of scattered to incident wave magnitude is not dependent on this quantity. However, when microwave devices operate in the nonlinear regime, these conditions no longer apply, and s-parameters cannot be used to capture the full behaviour of the device.

2.3 Measurements of Nonlinear Devices

Microwave devices operating in the nonlinear regime exhibit three differences from their linear counterparts which are significant to the designer:

1. The amplitude of electromagnetic waves scattered from the device are not linearly dependent on the amplitude of waves incident. This is the cause of features such as gain compression and gain expansion in amplifiers. Some of these effects are solely due to the nonlinear sources inside the device, while others are a symptom of the combined response of the nonlinearity and the power supply. A typical gain compression curve is shown in Fig. 4.

2. The frequency superposition principle does not apply, and instead the frequency spectrum of scattered waves contains components at frequencies other than those incident upon it. Rather than the incident signals purely summing inside the device, they are also multiplied

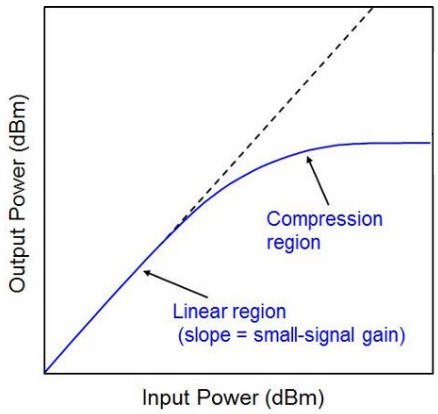


Figure 2.3: Gain compression occurs when an amplifier is driven into a nonlinear operating regime.

with each other (frequency mixing), as shown by

$$b = c_0 + c_1 a + c_2 a^2 + c_3 a^3 + \cdots, \tag{2.13}$$

$$\alpha = \beta = 2\pi\omega t, \tag{2.14}$$

$$a(t) = A\cos(\alpha), \tag{2.15}$$

$$\cos(\alpha)\cos(\beta) = \frac{1}{2}(\cos(\alpha+\beta) + cos(\alpha-\beta)), \tag{2.16}$$

$$a^2(t) = \frac{1}{2}A^2[\cos(2\pi(2\omega)t) + 1], \tag{2.17}$$

$$a^3(t) = \frac{1}{4}A^3[\cos(2\pi(3\omega)t) + 3\cos(2\pi\omega t)]. \tag{2.18}$$

Here, a and b are the incident and scattered waves for the device, c_i are coefficients of the device's nonlinear transfer function, and $a(t)$ is a wave in the time domain with amplitude A and frequency ω. For stimuli with a single frequency (α=β, as above), integer multiples of that frequency will be scattered from the device (harmonics). For stimuli with multiple tones ($\alpha \neq \beta$), additional products from combinations of the incident tone frequencies will be scattered (intermodulation). If the nonlinear device is incident with a fixed bandwidth of frequencies, such as the case for communications signals, then sidebands will be produced around the harmonics of the oscillator frequencies. This effect can be troublesome in practical designs where the unwanted sidebands overlap with the useful microwave bandwidth, distorting the signal. For this reason, it is important for designers to be able to accurately measure and characterise this nonlinear effect. Fig. 5 shows example spectra of these effects.

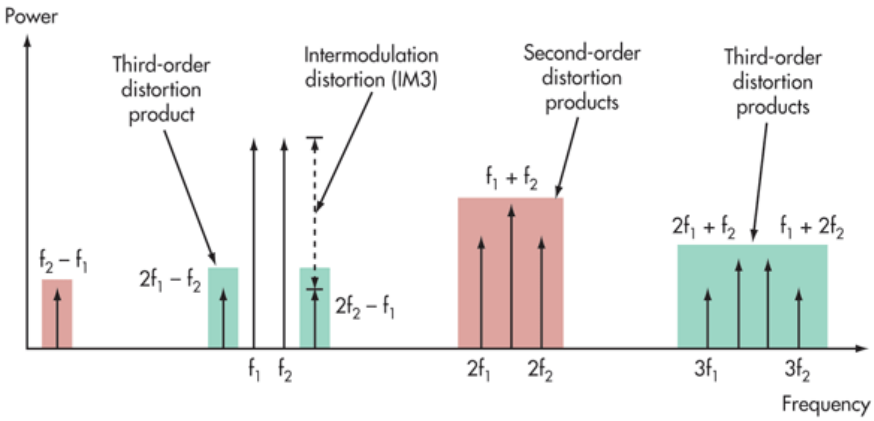


Figure 2.4: Intermodulation products from two tones within the cellular channel bandwidth f_1 and f_2. The second order products, and upper third order products, can be easily filtered out. However, the lower third order products $2f_1 - f_2$ and $2f_2 - f_1$ are located within the channel bandwidth and interact with the useful data, increasing EVM and BER **Hall2013**.

3. The amplitude of scattered waves with multiple incident waves is dependent (nonlinearly) on the phase of the incident waves. In the linear regime the superposition principle prevents this, but now there is a nonlinear dependence which can have significant effects on the amplitude of scattered waves. Designers must consider this when building efficient nonlinear amplifiers, which leads to the practice of accurately terminating scattered harmonic frequencies at an optimum phase. This will be covered in more detail in chapter X when we discuss nonlinear device models.

The result of these differences is that the measurement requirements for nonlinear devices are considerably larger than for linear devices. The nonlinear dependencies on stimulus power and phase means that ratioed measurements no longer fully capture the device response, and absolute measurements of the magnitude and phase of both the incident and scattered waves is required. The production of scattered waves at frequencies different to those in the stimulus demands an additional dimension of measurements. In contrast to These complications must be met with changes to both the measurement system and the method of storing the results.

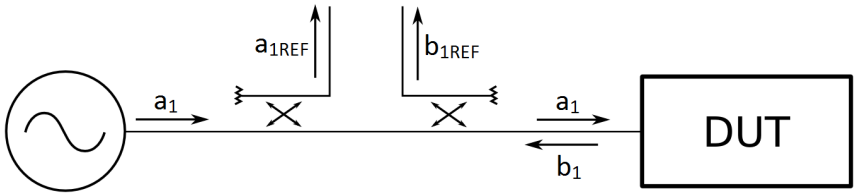


Figure 2.5: A one-port simple reflectometer. a_1 is the incident wave generated by the source, which is admitted to the DUT while also being sampled by the directional coupler and sent to the reference receiver via a_{1REF}. The reflected wave, b_1, is also sampled by another directional coupler and sent to the test receiver as b_1REF, with the remaining power dissipated at the matched source.

2.4 Vector Network Analysers

To measure the incident and scattered waves for a DUT and calculate the s-parameters as in (5), a vector network analyser (VNA) is typically used. The VNA is a quintessential piece of RF and microwave instrumentation and is found in most if not all such laboratories. Due to the challenging nature of measurements at these frequencies, it is a complicated instrument with many internal parts. This section explains how the VNA functions and the procedures behind its calibration. For a good history of VNA architecture and product development please refer to [Teppati Camb, Dunsmore Wiley].

2.4.1 Architecture

The origin of the VNA lies in an early instrument called a reflectometer. Designed in 1947 by Parzen and Yalow [x], it became an invaluable tool for characterising transmission lines used in telecommunication systems. Shown in Figure x, the incident signal is generated by a swept signal source and passes through the directional coupler before arriving at the DUT. The voltage reflection coefficient of the DUT will cause an amount of incident power to be reflected, which passes back through the coupler before being absorbed by the source (which has very low reflection). The directional couplers allow the waves travelling between the source and the DUT to be sampled by complex receivers, filtering the two waves by their direction of travel thus allowing the incident and scattered waves to be separated for measurement.

The limitation of a single reflectometer is that it can only measure waves at one port of

a DUT, therefore preventing transmission measurements. By adding a second reflectometer and synchronising the stimuli and measurements, it is possible to measure all s-parameters of a two-port device. This is the fundamental structure of a VNA, and most variations consist of changing the number of sources or receivers to optimise the instrument for cost or performance. Many older designs use an economical single source which is switched between both ports, whereas now the price of sources has falled, there are instruments available with two independent sources, which allows two-tone and some types of nonlinear measurements. These more versatile units often also expose more connections between internal components (e.g. the couplers and receivers) to allow the user to perform non-standard measurements or to add attenuation or preamplification for extreme stimulus powers. Modern VNAs also offer the option of measuring more than two ports, which are referred to as "multi-port" measurements. Several manufacturers offer four-port instruments which include four reflectometers (with usually two sources), although with external switching networks it is possible to expand this up to 48 ports [http://www.microwavejournal.com/articles/21785-vector-network-analysis-with-up-to-48-ports]. The basic block diagram of a modern two-port double-reflectometer VNA is shown in Fig. 9. To measure both stimulus conditions for the two-port S-parameter equations in (5), the sources alternate between delivering power and acting as a load for each measurement. As the source is swept the a and b waves for all ports are measured against frequency, from which the VNA software calculates the S-parameters. The receivers sampling the incident waves are known as the reference receivers and those sampling the scattered waves are called measurement or test receivers.

To perform S-parameter measurements using a VNA, the user must set both the frequency span and number of frequency points. They may also change settings of intermediate frequency bandwidth (IFBW) and numerical averaging, both of which reduce measurement noise by applying digital filtering but can consequently increase acquisition time. The user will then perform a calibration, which corrects for any response present in the measurement setup that is not caused by the DUT. When the system is calibrated physical 'measurement planes' are defined, where only effects of the signal path on the DUT side of the planes are incorporated in the measurement results. This is illustrated in Fig. 10. Once this step is complete, the VNA is ready for use. However, it is good practice to first check that calibration was successful by measuring some known devices (verification), or to use techniques such as ripple extraction (discussed in Chapter 4) to measure the residual uncertainty. This process characterises remaining error which the calibration failed to correct.

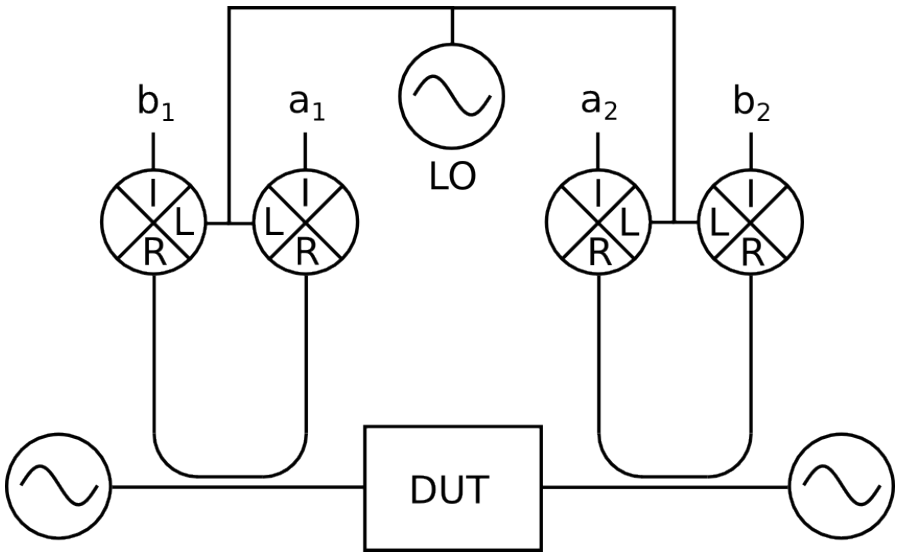


Figure 2.6: A modern two-source mixer-based VNA, which employs heterodyning to allow measurements at microwave frequencies. Two directional couplers are located between each source and the DUT and are connected back to back. These sample waves travelling in both directions and are connected to mixers which downconvert the microwave frequencies (R) into intermediate frequencies (I) which can be sampled by the complex receivers. The shared local oscillator (LO) feeding the mixers preserves phase coherence between the receivers. This configuration is known as a two-port double-reflectometer VNA. Figure adapted by author from **Root2013**.

2.4.2 Error Models

To remove the effect of the measurement setup from the measured device response, an error model is formed to capture the response of the measurement setup during calibration. These error models are stored in the memory of the VNA and are typically de-embedded from the measured device response before the results are presented to the user (although the raw measurements can still be obtained for separate post-processing). Because the measurement setup response is frequency dependent, the error model coefficients are characterised across the measurement bandwidth and are either applied at each measurement frequency or linearly interpolated.

2.4.2.1 One-Port Model

The classic one-port error model can be obtained through analysis of the signal flow diagram of a one-port VNA shown in Fig. 6. One can write the relationship between the measured (Γ_{M}) and absolute (Γ_{A}) reflection coefficients as

$$\Gamma_{\mathrm{M}} = D + \frac{T\Gamma_{\mathrm{A}}}{1 - M\Gamma_{\mathrm{A}}}, \tag{2.19}$$

where D, M and T are error coefficients which capture the unwanted response of the measurement setup. For this model, the three coefficients each have a physical meaning as they are caused by separate physical effects (illustrated in Fig. 7):

- **Directivity** (D) is caused by the nonideal operation of the directional couplers used to separate the incident and reflected waves inside the VNA. In practice, some amount of incident wave will travel into the test receiver port (vica versa??), reducing the measured gain of the device under test.
- **Test port match** (M) results from the impedance of the VNA test port (either the original test port or the extended measurement plane including any cables or other components in the setup) being different from the characteristic impedance of the measurement, which is typically 50-Ω. This effect will cause some of the incident wave to be reflected at the test port which is not due to the device response.
- **Reflection tracking** (T) characterises the insertion loss of the couplers and other measurement components between the reference receiver and the test receiver.

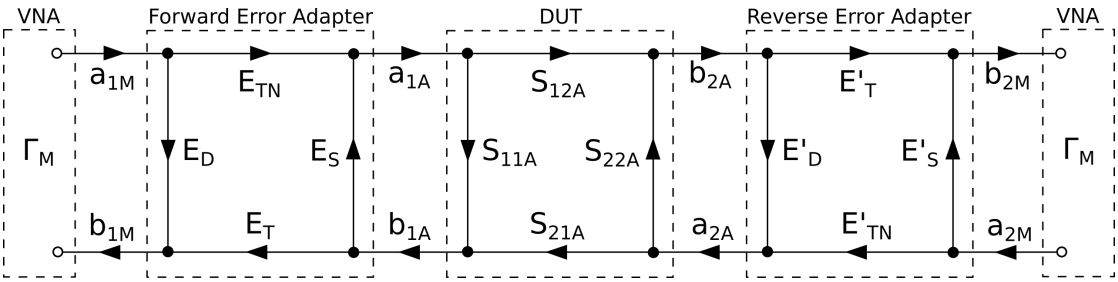


Figure 2.7: The 8-term error model for a two-port measurement. E_D, E_S, and E_T are the same as for the one-port model, except there are now sets of each for both ports. These extra terms account for different error values when the incident signal is sourced from each port. Additionally for the two-port case, a transmission term E_{TN} has been added for each direction.

2.4.2.2 8-Term Model

Devices with two or more ports require transmission measurements in addition to the reflection measurements which can be corrected using the one-port model. A popular two-port error model, the 8-term model, adds two transmission terms to the method used for the one-port model. This is shown in Fig. blah.

2.4.3 Calibration

2.5 Large Signal Network Analysers

2.5.1 Absolute 8-Term Error Model

2.5.2 Power Meter Calibration

2.5.3 Phase References

2.6 Conclusions

Testing, testing[1], [2].

Bibliography

[1] L. Stant, P. Aaen, and N. Ridler, "Evaluating residual errors in waveguide network analysers from microwave to submillimetre-wave frequencies," in *IET Colloquium on Millimetre-Wave and Terahertz Engineering & Technology 2016*, Institution of Engineering and Technology (IET), 2016. DOI: 10.1049/ic.2016.0016.

[2] ——, "Comparing methods for evaluating measurement uncertainty given in the JCGM 'evaluation of measurement data' documents," *Measurement*, vol. 94, pp. 847–851, Dec. 2016. DOI: 10.1016/j.measurement.2016.08.015.

3 Measurement Uncertainty

3.1 Introduction

A measurement is an observation of a physical effect or quantity which provides useful information. This information, through the ages, has been used to facilitate advancement of both scientific knowledge and industrial development - from the production of standardised stone blocks to build the pyramids of ancient Egypt, to the production of standardised car parts to build Henry Ford's Model T. In the scientific realm, advanced measurement techniques at laboratories such as CERN are used to convince the world that new subatomic particles exist.

To communicate information about a measurement, the recipient needs to be able to either make or imagine a similar observation to that of the original measurer (or metrologist). The simplest way of doing this is to provide the recipient with the same physical effect or quantity for which to make their own observation (if you require a new nut for a bolt from a hardware shop, you might intuitively take the bolt with you), however, this can be inconvenient or impractical with larger objects, or if the recipient is located far away. Instead, you might substitute a more portable representation. For example, if you were to measure the size of a doorway to see if a new piece of furniture may fit through it, you might cut a piece of string to the same length and use this as the representation of the width of the item. However, this approach is very wasteful and also impractical for many physical effects (temperature, flow, pressure).

A solution widely thought to have been first established in the 3rd or 4th Millennium BC (see Figure 3.1), is a system of units. In such a system, a discretised value of a quantity is standardised and knowledge of its value is disseminated to all people who wish to use it. Typically, a range of discrete values are chosen, such that the system of units can be conveniently used to represent all measurements. Knowledge of the discretised values is obtained from a primary standard which becomes the definition of the unit and is used to create copies of the standard which can be

Figure 3.1: Egyptian royal cubit rod of Maya (treasurer of King Tutankhamun) 1336–1327 BC. The cubit is thought to be the earliest attested standard measure of length, first used in the 3rd or 4th Millennium BC.

given to users of the unit system to perform measurements with. The most common method of performing measurements with a unit system is to use a standard to calibrate a measuring instrument, which can then be used to measure an arbitrary value of a quantity in the units defined by the standard.

The introduction of a regulated system of units enables commerce, as traded goods can be reliably valued between merchants across cities. This application is encountered by all citizens, and so there is a high demand for standards to be produced from the primary standard and physically distributed. It becomes impractical to create all standards by copying the primary standard directly (in some cases because the value of the primary standard is perturbed each time it is measured), and so a tiered organisational structure of standards is used. In this structure, there is a tier consisting of a small number of standards which are created directly from measurements of the primary standard, followed by subsequent tiers of larger numbers of standards which are derived from measurements of those in the previous tier. For any standard produced, it should be possible to trace the lineage back to a measurement of the primary standard. This is referred to as a traceability chain (see Figure 3.2) and it is a fundamental tenet of metrology. Measurements with a shorter traceability chain are considered more traceable than those with longer chains.

Today, the primary standards are maintained in most countries by a National Measurement Institute (NMI) and co-ordinated by the Bureau of International Weights and Measures (BIPM). To accommodate international trade and compatibility, a routine process of inter-comparisons is undertaken to ensure that the values of the primary standards between countries are in agreement.

Secondary standards are also kept by the NMIs and are used to reduce excessive wear to

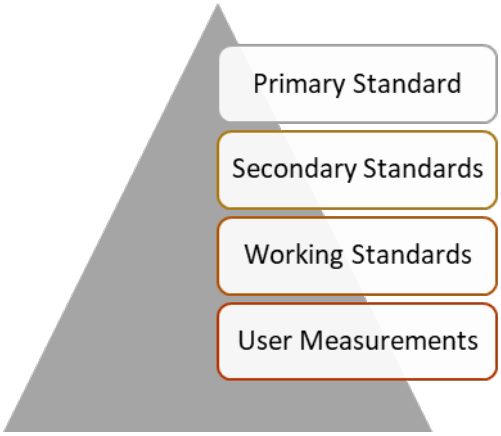


Figure 3.2: The traceability chain, where the pyramid shows the number of instances of standards in each tier. Secondary standards are held at NMIs and used to periodically calibrate working standards, which are sent by manufacturers and laboratories. User measurements are made using instruments calibrated with these working standards, so they number the greatest and are at the bottom of the traceability chain.

the primary standard caused by frequent measurements (and also to reduce bottlenecks caused by having a single standard). They are calibrated against the primary standard as infrequently as possible, again to reduce wear. Secondary standards are used by the NMI to characterise working standards which are sent to them by manufacturers and research institutes. Another important task of each NMI is to perform investigations to discover new and improved methods of measurement, which make use of secondary standards to better compare the accuracy of different methods.

Working standards are used, for example, by instrumentation manufacturers who may use them to calibrate their products before shipping to the customer, and more generally the standards can be used to calibrate test equipment to identify faulty products. Larger research institutes typically use working standards to recalibrate instrumentation prior to performing very sensitive measurements. To ensure that product specifications and scientific measurements are traceable and of high quality, accreditation services such as the United Kingdom Accreditation Service (UKAS) exist to certify manufacturers and laboratories that demonstrate good measurement practice and use traceable measurements [1].

The selection of quantities for which primary standards are kept is only a subset of those for which recognised units exist. This is because many units are derived quantities, where

their value can be obtained by calculation using definitions of other units. For example, the definition of the unit of resistance (R, ohms) can be derived from that of voltage (V, volts) and current (I, amperes), because $R = V/I$. The eight fundamental "base" units which make up the International System of Units (SI), are the metre, kilogram, second, ampere, kelvin, candela and mole. From these unit definitions, it is possible to define any other derived unit in use. NMIs will usually keep secondary standards of most derived quantities that users may wish to calibrate against, which are traceable to one or more primary standards of the base units. Although traditionally all primary standards were defined by physical artefacts (e.g. metallic weights, burning candles), these are being gradually replaced by definitions involving physical constants (e.g. Plank, Boltzmann), which do not degrade over time or use. The "Ninth SI Units" [2], a proposition recently accepted by the BIPM, covers the redefinition of four of the SI units (the ampere, the kilogram, the kelvin and the mole) which is scheduled for May 2019.

The crucial effect of traceability on measurements is the confidence in their results. Measurements with poor traceability (longer chains) will produce results which are likely to be less accurate than those with better traceability (shorter chains). The reason for this is measurement uncertainty, which will now be explained.

It is impossible to know the true value of a quantity being measured as many undesirable physical effects typically occur during the measurement process. These effects contribute error (an unwanted perturbation) to the measured value, causing a reduction in accuracy (the deviation of the measured value from the true value). Typical sources of error in measurement include thermal noise, imperfect calibration and drift of environmental conditions from those at which a measuring instrument was calibrated. In some cases, it is possible to quantify and correct for these errors, but there are often many sources (some of which contribute very small errors) which cannot be corrected for. This is because either the error cannot be quantified or the value of the error will change over the duration of the measurement process (random errors). Any source of error which cannot be removed from a measurement becomes a source of uncertainty, because the deviation of the measured value from the true value due to this source of error is uncertain. If it is possible to quantify the amount of uncertainty in a measurement, then a degree of confidence can be formed about its value. If every measurement has an associated uncertainty in its value, then any measurement involving the results of previous measurements will include uncertainty contributions from both measurements. Measurements with good traceability involve fewer sources of uncertainty than those with poor traceability, leading to a higher degree of confidence in the former. It is because of this fact that NMIs strive to reduce the uncertainties

in their primary standard definitions, which in turn reduces the uncertainty in all traceable measurements.

Because it is impossible to know the amount of error in a source of uncertainty, probability and statistical theories are used to instead describe the amount of uncertainty associated with it. By the nature of these theories there are often several methods which can be used to obtain a result, which sometimes provide different values. To ensure consistency and portability of uncertainty definitions, measurement guides were created in each industry and area of science, which specialised in processing the results of typical measurements. In addition, different guides were produced depending on the level of accuracy required - as more accurate measurements often require more effort to complete. Although this practice allowed suitable measurement comparisons within each field (e.g. chemistry, mechanical engineering), ambiguities still existed in uncertainty definitions between fields. To address this, a landmark document was published in 1993 by the International Organisation for Standardisation (ISO), the Guide to the Expression of Uncertainty in Measurement (GUM) [3]. This document was the work of representatives from seven international organisations: the BIPM, the International Organisation of Legal Metrology (OIML), the International Electrotechnical Commission (IEC), the ISO, the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), the International Union of Pure and Applied Chemistry (IUPAC), and the International Union of Pure and Applied Physics (IUPAP). The GUM, updated in 2008 [4], is still used today as a reference for the evaluation of measurement uncertainty in many laboratories and industries across the world. The seven original organisations which wrote the GUM, together with the International Laboratory Accreditation Cooperation (ILAC, of which UKAS is a member), form the Joint Committee for Guides in Metrology (JCGM), who maintain the GUM and subsequent additional documents. These additional documents consist of the International Vocabulary of Metrology (VIM) [5] and two supplements to the GUM [6], [7]: Supplement 1 covers the use of a Monte Carlo method [8] in uncertainty evaluation; Supplement 2 is used where more than one quantity is measured at the same time (multivariate).

Throughout this dissertation, the methodologies presented in the GUM will be used. The international authority of the guide, developed by seven international organisations (including the two global standardisation bodies IEC and ISO), gives strong motivation to use it as a basis for a framework to evaluate uncertainty in measurement.

This Chapter describes the evaluation of uncertainty prescribed in the GUM and highlights an inconsistency in the current version of the GUM and associated documents (which can have

a profound effect on electromagnetic measurements).

3.2 The Measurement Process

In contrast to basic evaluations of uncertainty, where only repeat measurements of the quantity of interest are analysed, the GUM prescribes a more rigorous approach, which defines a mathematical model of the measurement process (measurement model) and propagates uncertainty through that model to the result (measurands). This allows any uncertainties from previous measurements, including those involving standards in the traceability chain, to be included in the result. The measurement model can be simple, such as measuring resistance using input quantities of voltage and current, or complicated and multivariate, requiring many input quantities and producing many output quantities. In some cases, the measurement model may not be known and can be defined as a black box, but this has certain limitations discussed later with Monte Carlo methods.

The GUM defines a process that is to be followed when evaluating uncertainty in measurement. It consists of the following steps:

1. Modelling the measurement.
2. Evaluating standard uncertainty of input quantities.
3. Determining combined standard uncertainty of the measurands.
4. Determining expanded uncertainty of the measurands.

where standard uncertainty is an uncertainty expressed as a standard deviation and expanded uncertainty is used to define a coverage interval encompassing a large fraction of the distribution of values that could reasonably be attributed to the measurand.

3.2.1 Modelling the Measurement

The VIM document [5] defines a measurement model[1] as "a mathematical relation among all quantities known to be involved in a measurement". In many cases, where an explicit relation can be written, it is possible to further define a measurement function. We can represent this

[1]The definition of model used in "measurement model" is different to that used when describing models of amplifiers seen elsewhere in this dissertation.

generally as a set of measurands Y having a functional relationship, $f(.)$, depending on N input quantities $X_1, X_2, \ldots, X_N$:

$$Y = f(X_1, X_2, \ldots, X_N) \tag{3.1}$$

The estimates of the measurands $\bar{Y}$ can be found by evaluating the measurement model using the estimates of each input quantity $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N$:

$$\bar{Y} = f(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_N) \tag{3.2}$$

Each input quantity could either be observed during the present measurement, a result from a previous measurement, or another source of information such as a datasheet or specification. An example of a measurement model could be for a temperature measurement, where the input quantities would include the value observed from the meter, the previously measured values of two calibration temperatures, and the assumed values of those calibration temperatures. Using this method, uncertainty from the calibration can be included in the evaluation. This is especially true for uncertainties caused by systematic errors, which do not vary during the measurement process and cannot be evaluated purely by performing repeat measurements.

3.2.2 Evaluating Standard Uncertainty of Input Quantities

Sources, or components, of uncertainty in measurement can be divided into two categories: Category A uncertainty components are those that are evaluated using statistical analysis of a series of observations (i.e. repeats); Category B components are those that are evaluated using other means, for example using information from datasheets.

The GUM presents methods that include the use of both Bayesian and classical probabilistic methods to evaluate the uncertainty in the input quantities for a measurement model. In particular, classical methods [9] are used for the treatment of Category A uncertainty components and Bayesian methods [10] are used for the treatment of Category B uncertainty components. This is a sensible assignment as classical (frequentist) methods work well for repeat observations and Bayesian inference can be used to incorporate alternative sources of knowledge. An informative discussion on these types of method can be found in [11]. Since the publication of the GUM, some authors have stated (for example, in [12]–[15]) that this combination of different probabilistic methods (i.e., Bayesian and classical) represents an inconsistency in the GUM methodology for evaluating measurement uncertainty. The author has published a paper considering the effects of

this inconsistency on electromagnetic wave measurements at radio frequencies [16], which forms the basis for this section of the chapter.

The supplements to the GUM [6], [7] resolve the above-mentioned inconsistency by introducing a method for treating the Category A uncertainties that follows a Bayesian approach [17]. Therefore, the two supplements no longer contain the inconsistency found in the original GUM document. However, as a consequence of this change, there is now inconsistency between the method used to evaluate uncertainty described in the GUM and that described in the two supplements. In many situations, these different methods do not have a significant impact on the overall uncertainty that is evaluated. For situations where a considerable number of input quantities are observed simultaneously, the two different approaches can produce significantly different values of uncertainty. Such situations often occur in the area of high-frequency electromagnetic metrology, which is the topic of this dissertation.

3.2.2.1 Category A Evaluation

GUM Method The classical statistical technique [9] applied to Category A uncertainties in the current GUM is based on a series of observations of a randomly varying input quantity. After n observations $x_1, x_2, \ldots, x_n$, the best available estimate (arithmetic mean of measured values), $\bar{x}$, and standard deviation, s, of a randomly varying input quantity, X, is written as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{3.3}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3.4}$$

respectively, where x_i is the result of the ith observation. Importantly, a minimum of two observations must be made ($n = 2$) in order for $\bar{x}$ and s to be defined. The standard uncertainty of the best estimate of X, $u(\bar{x})_{\mathrm{GUM}}$ can be found by dividing s by the square root of the number of observations:

$$u(\bar{x})_{\mathrm{GUM}} = \frac{s}{\sqrt{n}} \tag{3.5}$$

If there are correlated (mutually dependent) input quantities present in the measurement model, the covariances of each pair of input quantities must also be calculated before the propagation of uncertainty. Both the standard uncertainties and the covariances for N input quantities

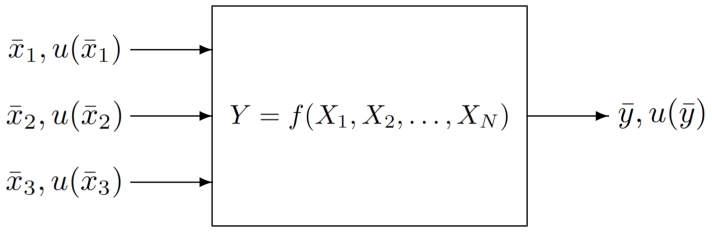


Figure 3.3: The propagation of measurement uncertainty through the measurement model, as specified in the GUM. Here, a single measurand Y is shown, but the model could include multiple measurands.

can be represented in a symmetric $(N \times N)$ matrix containing the variance of each quantity (s^2) along the diagonal and the covariance between x_i and x_j in the i, jth element. This is called the "uncertainty matrix" in the GUM and the "measurement covariance matrix" in the GUM Supplement 2. An example given in the GUM and described later in this Chapter, demonstrates this scenario using the example of a simultaneous measurement of resistance and reactance with voltage, current and phase as correlated input quantities [4, Example H.2].

Once the uncertainties of the input quantities have been evaluated, they are propagated through the measurement model, as illustrated in Figure 3.3. This requires the sensitivities of the measurand to each input quantity to be calculated to at least a first order approximation (i.e. the partial derivatives of the measurement model known). The estimates of the input quantities are used in the measurement model to obtain the estimate of the measurand. The variances and covariances of the input quantities are combined with the sensitivity coefficients in order to obtain the variance of the measurand. The combined standard uncertainty of the measurand is equal to the positive square root of this value. The result of the measurement process is then presented as the measurand estimate and combined standard uncertainty. Alternatively, the combined standard uncertainty can multiplied by a positive factor to form an expanded measurement uncertainty. From this value, a coverage interval can be derived which states a particular probability that the true value of the measurand is within that range. A more detailed description of propagating uncertainties through the measurement model and presenting the results will be given later in this Chapter, but a brief introduction was given here to help the reader understand research on input quantity uncertainties presented in the remainder of this section.

GUM Supplement Method Both GUM supplements (GUM-S1/S2) [6], [7] use a Bayesian approach [18] to assign a probability density function (PDF) to all input quantities. This approach results in the choice of a t-distribution to characterize Category A input quantities, in contrast to the Gaussian distribution used in the GUM [7, para. 5.3.2.1]. Of particular relevance is the inclusion of the degrees-of-freedom parameter, ν, in the definition of the standard uncertainty and covariances of a t-distribution. Whereas for the Gaussian distribution ν is used as a measure of reliability of the standard uncertainty, it is explicitly required when using the t-distribution in order to obtain the standard uncertainty, $u(\bar{x})_{\mathrm{SUPP}}$:

$$u(\bar{x})_{\mathrm{SUPP}} = \frac{s}{\sqrt{n}} \times \sqrt{\frac{\nu}{\nu - 2}}, \tag{3.6}$$

where $\nu = n - N$, with n being the number of observations and N being the number of input quantities. In the GUM-S1 only a univariate t-distribution is offered, which represents $N = 1$ input quantities. For this case (3.6) can be rewritten as:

$$u(\bar{x})_{\mathrm{SUPP}} = \frac{s}{\sqrt{n}} \times \sqrt{\frac{n-1}{n-3}}. \tag{3.7}$$

Equation 3.7 is undefined if n is less than four, in which case the standard uncertainty cannot be calculated for a single input quantity according to the guidance given in the GUM-S1 (and the GUM-S2). Figure 3.4 illustrates the ratio between the standard uncertainty values calculated for different numbers of observations of a single Category A input quantity using the GUM and the GUM-S1/S2 approaches. It can be seen that when $n = 4$, $u(\bar{x})_{\mathrm{SUPP}} = \sqrt{3} \times u(\bar{x})_{\mathrm{GUM}}$, and as the number of observations increases the results from both approaches converge: If n tends to infinity, the t-distribution tends towards a Gaussian distribution. However, most commercial laboratories would avoid making large numbers of measurements as this is often time-consuming and therefore expensive.

For measurements involving multiple input quantities, such as the measurement of a vector quantity, a multivariate/joint distribution should be used as suggested in the GUM-S2. The variances and covariances between all pairs of input quantities are obtained using a matrix form of (3.6) ([7, Section 5.3.2]):

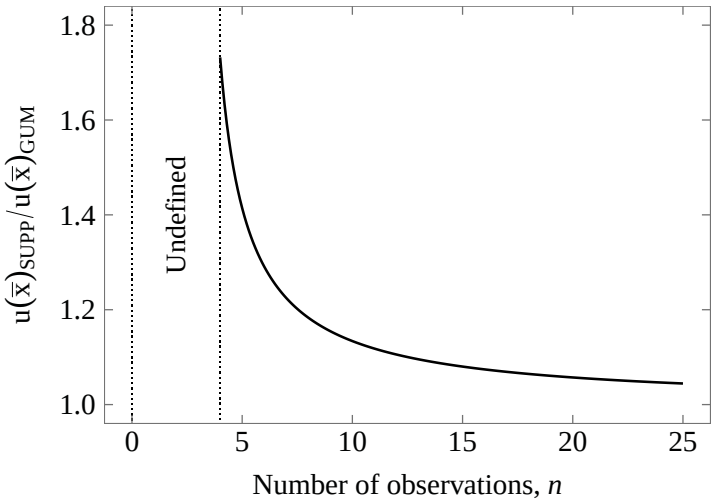


Figure 3.4: Scaling factor to convert from a GUM standard uncertainty to a GUM Supplement.

$$\boldsymbol{V}(\boldsymbol{X}) = \frac{\nu}{(\nu-2)}\frac{\boldsymbol{S}(\boldsymbol{X})}{n} = \frac{1}{n(n-N-2)}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^{\top} \tag{3.8}$$

$$\boldsymbol{S}(\boldsymbol{X}) = \frac{1}{\nu}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^{\top} \tag{3.9}$$

$$\boldsymbol{V}(\boldsymbol{X}) = \begin{bmatrix} u(\boldsymbol{x}_1)^2 & u(\boldsymbol{x}_1,\boldsymbol{x}_2) & \dots & u(\boldsymbol{x}_1,\boldsymbol{x}_n) \\ u(\boldsymbol{x}_2,\boldsymbol{x}_1) & u(\boldsymbol{x}_2)^2 & \dots & u(\boldsymbol{x}_2,\boldsymbol{x}_n) \\ \vdots & \vdots & \ddots & \vdots \\ u(\boldsymbol{x}_n,\boldsymbol{x}_1) & u(\boldsymbol{x}_n,\boldsymbol{x}_2) & \dots & u(\boldsymbol{x}_n)^2 \end{bmatrix} \tag{3.10}$$

where $\boldsymbol{U}_{\boldsymbol{X}}$ is the uncertainty matrix, $\boldsymbol{x}_i$ is a sample from the array of vectors containing input quantity indications and $\bar{\boldsymbol{x}}$ is the arithmetic mean of that array. For this multivariate case, the minimum value of n will increase linearly with N, such that the standard uncertainty is undefined unless $n > N + 2$.

Comparison of GUM and GUM Supplements approach using example H.2/9.4 Both the GUM and the GUM-S2 provide an identical example which can be used to compare the different standard uncertainties. The example is a simultaneous measurement of resistance and reactance, which uses a measurement model with multiple input quantities and multiple output quantities (measurands). The input quantities are voltage V, current, I, and phase, ϕ, and the measurands are resistance R, reactance, X, and impedance, Z. The measurement model is

defined as:

$$R = \frac{V}{I}\cos\theta, \quad X = \frac{V}{I}\sin\theta, \quad Z = \frac{V}{I} \tag{3.11}$$

Six sets of observations [5] ($n = 6$) of V; I; ϕ are obtained independently by measurement. The version of this example given in the GUM uses only $n = 5$ sets, but one additional set of values of V; I; ϕ has been added for the GUM-S2 example to allow (3.6) to be defined for $N = 3$ input quantities, a condition which was explained at the end of the previous section. These values, together with their arithmetic means and standard uncertainties as calculated from the two approaches using (3.5) and the matrix form of (3.6) (which is applicable to measurements involving multiple input quantities), are presented in Table 3.1. The ratios of the standard uncertainties from each approach is also included in the table, which are identical for all these input quantities due to their dependence only on n and N, which are also equal for all these input quantities (e.g. when $n = 6$ and $N = 3$, $\sqrt{(\nu/(\nu-2))} = \sqrt{((n-N)/(n-N-2))} = \sqrt{3}$. This explains why standard uncertainties evaluated with Category A methods using the minimum number of observations following the GUM-S1/S2 approach are always 1.732 times larger than the standard uncertainties calculated following the GUM approach.

This difference in the input quantity uncertainties calculated from the two approaches propagates through the measurement model and therefore significantly affects the combined standard uncertainties of the measurands. Table 3.2 presents the combined standard uncertainties of the measurands for the described example as evaluated by both approaches, together with a ratio of the uncertainty values. For all three measurands the combined standard uncertainty calculated using the GUM-S1/S2 method is more than double the equivalent values calculated using the GUM method. For other measurement models with higher sensitivities to the input quantities, this difference could be even greater.

Comparison of GUM and GUM Supplements approach using microwave scattering parameters example High-frequency electromagnetic metrology often involves using multiple complex-valued quantities. Common input quantities for this type of measurement, measured using instruments such as vector network analysers (VNA), are scattering parameters (S-parameters), as described in Chapter 2. Because each S-parameter is a complex-valued quantity ($S = (S_{\mathrm{Re}}, S_{\mathrm{Im}})$), there are $2m^2$ input quantities required in a measurement model for the complete response of an m port device. All these quantities are correlated, so a multivariate distribution should be used to represent them. It has been shown previously that for a Category

Value	V/V	I/A	ϕ/rad
x_1	5.007	19.663	1.0456
x_2	4.994	19.639	1.0438
x_3	5.005	19.640	1.0468
x_4	4.990	19.685	1.0428
x_5	4.999	19.678	1.0433
x_6	4.999	19.661	1.0445
$\bar{x}$	4.9990	19.6610	1.04446
$u(\bar{x})$GUM	0.0026	0.0077	0.00061
$u(\bar{x})$SUPP	0.0045	0.0134	0.0011
$\frac{u(\bar{x})\text{GUM}}{u(\bar{x})\text{SUPP}}$	1.732	1.732	1.732

Table 3.1: The indication values from the example "Simultaneous Resistance and Reactance Measurement" and their statistical properties as evaluated by the approaches given in [4, Example H.2] and [7, Example 9.4].

Method	$u(R)/\Omega$	$u(X)/\Omega$	$u(Z)/\Omega$
GUM	0.058	0.241	0.193
GUM-S2	0.130	0.540	0.431
$\frac{\text{GUM-S2}}{\text{GUM}}$	2.241	2.241	2.233

Table 3.2: A comparison of the results obtained for the example "Simultaneous Resistance and Reactance Measurement" using the approaches given in [4, Example H.2] and [7, Example 9.4].

Ports, m	Input quantities, N	Required minimum number of repeat observations, n, for $u(\bar{x})_{SUPP}$ to be defined	$\frac{u(\bar{x})\text{GUM}}{u(\bar{x})\text{SUPP}}$
1	2	5	1.732
2	8	11	1.732
3	18	21	1.732
4	32	35	1.732
⋮	⋮	⋮	1.732
8	128	131	1.732

Table 3.3: The difference in standard uncertainties obtained using the GUM ($u(\bar{x})_{GUM}$) and the GUM-S1/S2 ($u(\bar{x})_{SUPP}$) approaches to measure a full set of scattering parameters for microwave devices with various numbers of ports, m. Each device has $2m^2$ input quantities, N, and requires a minimum of $N + 3$ repeat observations, n, in order for $u(\bar{x})_{SUPP}$ to be defined.

A evaluation of uncertainty, both the number of repeat observations and the number of input quantities have a significant effect on the difference in uncertainty as calculated from the two approaches presented in the GUM and the GUM-S1/S2. Table 3.3 shows the ratio of uncertainties calculated from both approaches when applied to a measurement using scattering parameters obtained from the minimum number of repeat observations, n, for devices with m ports.

It can be seen that for devices with multiple ports, n can become large in order for 3.6 to be defined and calculate the standard uncertainty. It is often the case that the user will not always have the time or resources available to perform such a quantity of measurements. In microwave measurement environments, connections are typically made by hand using coaxial connectors. A typical measurement may include a Category A evaluation of uncertainty due to connection repeatability. Considering the specific example of a 4-port device, this requirement would result in the need for a minimum of $35 \times 4 = 140$ repeat coaxial connections to be made in order to perform a Category A evaluation of the standard uncertainty using the GUM-S1/S2 approach. By contrast, the classical approach used in the GUM is defined with just 2 repeat observations, which would require only $2 \times 4 = 8$ repeat coaxial connections to be made. Figure 3.5 shows the minimum number of repeat observations required when using the GUM-S1/S2 approach, n, in order to be able to calculate a Category A evaluation of the standard uncertainty of a full

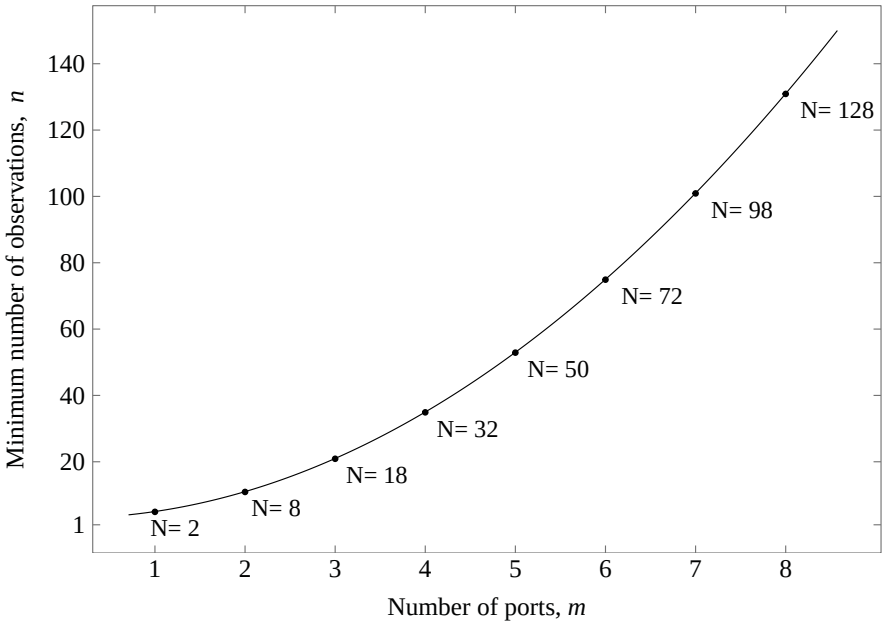


Figure 3.5: The minimum number of observations, n, required to calculate the standard uncertainty of a full set of S-parameters for a microwave device with m ports using the GUM-S2 approach. The number of input quantities, N, for each device is also shown.

set of S-parameters for a microwave device with m ports. In all cases, the standard uncertainty obtained using the GUM-S1/S2 approach is approximately 1.7 times larger than that obtained using the GUM approach.

Discussion The inconsistency of the approaches used in the GUM and its supplements to calculate the standard uncertainty of Category A input quantities of a measurement has two noticeable consequences:

1. There can be a large difference in the standard uncertainties reported by each approach. It is not straightforward to decide which is the correct approach to use, however the GUM approach is likely to be more attractive to commercial laboratories and test engineers since this leads to achieving smaller uncertainties in their results.

2. For situations involving multiple Category A input quantities, the Bayesian approach introduced in the GUM-S1/S2 can require a large number of observations before the standard uncertainty can be defined. Although the standard uncertainty calculated using the GUM approach will become less reliable with fewer observations, it is still possible to obtain

a result with only two observations of any number of input quantities. In a commercial laboratory the additional measurements required by the GUM-S1/S2 approach can be impractical, with many laboratories typically using only two or three measurements per device following the GUM approach. For a single input quantity this would require a potential doubling of the number of observations and therefore the test duration, which would either slow throughput or require more test stations to be added. If implemented, the additional time or financial investment would then produce uncertainties that are significantly larger than those obtained using the GUM approach.

This inconsistency is yet to be resolved, and the draft of an updated GUM which replaced much of the remaining classical approach with Bayesian techniques received many poor reviews when circulated for discussion. Work is now being carried out to find solutions to the issues raised by converting to a fully Bayesian GUM. Specific to the example presented in this Section, an article was recently published which offers a way to use Bayesian statistics to evaluate uncertainty in Category A input quantities with $n \geq 2$ repeat observations, which is the same number required by the classical approach [19].

For the work in this thesis, which is based on multivariate electromagnetic measurement problems, the GUM approach (instead of the Supplement 2 approach) is used. In addition, an existing software framework, introduced later, which is included as part of the complete framework presented in this work, also uses the GUM approach for processing Category A uncertainty components.

3.2.2.2 Category B Evaluation

Category B uncertainty components are those which have not been obtained by repeated measurements. Possible sources include previous measurement data, experience or knowledge of relevant materials and instruments, manufacturer's specifications, data provided by calibration and other certificates and reference data from handbooks.

Values obtained from these sources will typically be an estimate accompanied by either a standard uncertainty or an expanded uncertainty. The latter can be converted to a standard uncertainty, the process of which is described in Section 3.6. Category B uncertainty components are not restricted to Gaussian or t-distributions, and could for example be normal (rectangular), beta, or Cauchy distributions. Unless the combined standard uncertainty is determined via a Monte Carlo method, as explained in the following section, the standard uncertainty must be

known for the value to be used as an input quantity.

3.2.3 Evaluating Combined Standard Uncertainty

In order to determine the standard uncertainty of the measurand (the combined standard uncertainty) the uncertainties of the input quantities must be propagated through the measurement model. The GUM offers several methods to achieve this, which will be described in this section.

3.2.3.1 Monte Carlo Methods

Supplement 1 of the GUM [6] covers the use of a Monte Carlo technique to determine combined standard uncertainty in the measurand. The Monte Carlo technique has three important benefits for the propagation of uncertainty:

1. The measurement model does not need to be known explicitly. In some cases, the algorithm used to obtain a measurement result is proprietary and cannot be made available to the metrologist. Alternatively, the measurement model may be very complicated or involve numerical solutions which cannot be differentiated as required by other propagation methods.

2. Full knowledge of the probability distributions of the input quantities are used and preserved through the uncertainty propagation. Because the input quantity distributions are sampled directly, the complete probability distribution of the measurand can be obtained (see Figure 3.6). This can be very useful when more exotic distributions such as u-shaped distributions are used for input quantities, or if the measurement model is strongly nonlinear, when one cannot make assumptions about the probability distribution of the measurand.

3. The uncertainty propagation preserves nonlinearities in the measurement model. Alternative propagation methods presented in the GUM cause the measurement model to be linearised around the estimate. In most cases where a nonlinear measurement model is used, however, the uncertainty values are sufficiently small that a linear approximation is valid [4, p. 5.1.5]. Often, an initial Monte Carlo propagation is used to validate this assumption.

4. All correlations between input quantities are preserved. For many measurements involving multiple input quantities (especially in electromagnetic measurements), the uncertainties

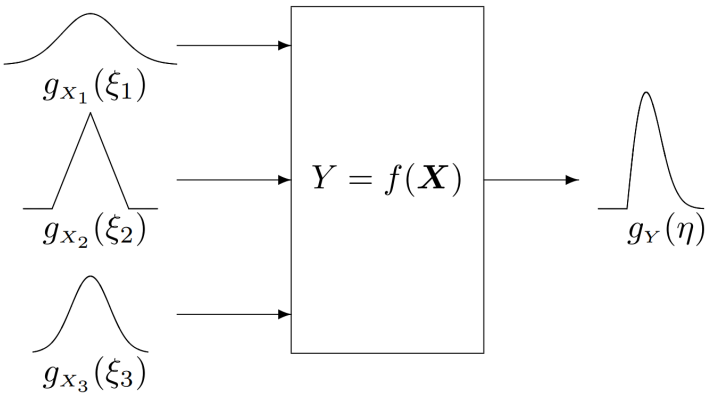


Figure 3.6: An illustration of the propagation of distributions from three input quantities g_{X1}, g_{X2}, g_{X3}, through the measurement model, Y, to the measurand, g_Y [6].

of one or more input quantities may be correlated. This means that when the value of one quantity changes, it affects the values of others. This can both increase or decrease the combined standard uncertainty in the measurand significantly. Chapter 4 will discuss the impact of correlations on VNA measurements.

The primary disadvantage of Monte Carlo methods is the time required to process them. For an accurate evaluation of uncertainty, the number of samples must be sufficiently large. Generally, the GUM recommends 106 samples for a 95% coverage interval accurate to one or two significant digits [6, p. 7.2.1]. The number of samples increases with the size of the desired coverage interval of similar accuracy. For many measurements today, the processing power of modern computers is sufficient for the duration of uncertainty propagations using Monte Carlo methods to be acceptable. However, in situations where the measurement model is very time-consuming to process, or where the uncertainty evaluation must be very fast, linear propagation techniques may be preferred. A detailed explanation of the steps involved in performing a Monte Carlo propagation can be found in Section 7 of [6].

3.2.3.2 Law of Propagation of Uncertainty

The primary propagation method presented in the GUM is the Law of Propagation of Uncertainty (LPU). This method uses first-order derivatives of the measurement model, together with the variances (and co-variances) of the input quantities, to determine a value for the combined standard uncertainty. The use of first-order derivatives means that the measurement model is

linearised, which in many applications can be a valid assumption. The LPU provides different equations for combining independent (uncorrelated) and correlated input quantities. For independent input quantities,

$$u_c^2(\bar{y}) = \sum_{i=1}^{N} (\frac{\partial f}{\partial x_i})^2 u^2(\bar{x}_i), \tag{3.12}$$

where $u_c^2(\bar{y})$, the combined variance, is the square of the combined standard uncertainty, f is the function describing the measurement model, and x_i is an input quantity with variance $u^2(\bar{x}_i)$. For correlated input quantities (whose covariances always form a symmetric and positive semi-definite matrix),

$$u_c^2(\bar{y}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} u(\bar{x}_i, \bar{x}_j) = \sum_{i=1}^{N} (\frac{\partial f}{\partial x_i})^2 u^2(\bar{x}_i) + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \frac{\partial f}{\partial x_i}\frac{\partial f}{\partial x_j} u(\bar{x}_i, \bar{x}_j). \tag{3.13}$$

Supplement 2 to the GUM offers a matrix formulation of (3.13) [7, p. 6.2.1.3], which handles the multivariate case where multiple measurands are encountered. If the covariance matrix of dimension $N \times N$ associated with $\bar{\boldsymbol{x}}$ is

$$\boldsymbol{U}(\bar{\boldsymbol{x}}) = \begin{bmatrix} u(\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_1) & \dots & u(\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_n) \\ \vdots & \ddots & \vdots \\ u(\bar{\boldsymbol{x}}_n \bar{\boldsymbol{x}}_1) & \dots & u(\bar{\boldsymbol{x}}_n, \bar{\boldsymbol{x}}_n) \end{bmatrix} \tag{3.14}$$

the covariance matrix of dimension $m \times m$ associated with $\bar{y}$ is

$$\boldsymbol{U}(\bar{\boldsymbol{y}}) = \begin{bmatrix} u(\bar{\boldsymbol{y}}_1, \bar{\boldsymbol{y}}_1) & \dots & u(\bar{\boldsymbol{y}}_1, \bar{\boldsymbol{y}}_n) \\ \vdots & \ddots & \vdots \\ u(\bar{\boldsymbol{y}}_n \bar{\boldsymbol{y}}_1) & \dots & u(\bar{\boldsymbol{y}}_n, \bar{\boldsymbol{y}}_n) \end{bmatrix} \tag{3.15}$$

and the sensitivity matrix $\boldsymbol{C}_{\bar{\boldsymbol{x}}}$ of dimension $m \times N$ containing the first-order partial derivatives of the measurement model to each input quantity (the Jacobian of the measurement model) is given by evaluating

$$\boldsymbol{C}_{\bar{\boldsymbol{x}}} = \begin{bmatrix} \frac{\partial f_1}{\partial X_1} & \cdots & \frac{\partial f_1}{\partial X_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial X_1} & \cdots & \frac{\partial f_m}{\partial X_N} \end{bmatrix} \tag{3.16}$$

at $\boldsymbol{X} = \bar{\boldsymbol{x}}$, then $\boldsymbol{U}_{\bar{\boldsymbol{y}}}$ is given by

$$\boldsymbol{U}_{\bar{\boldsymbol{y}}} = \boldsymbol{C}_{\bar{\boldsymbol{x}}} \boldsymbol{U}_{\bar{\boldsymbol{x}}} \boldsymbol{C}_{\bar{\boldsymbol{x}}}^{\top} \tag{3.17}$$

The LPU does not provide any information about the shape of the probability distribution of $\boldsymbol{U}_{\bar{\boldsymbol{y}}}$ or its components. The results of the measurement are obtained from the estimates and the combined standard uncertainties – the positive square roots of the diagonal terms of the covariance matrix.

3.2.3.3 Finite Difference Methods

Included in the definition of the LPU is an alternative method of determining the sensitivity coefficients in (3.16), without the need to know the measurement model, f, explicitly. This technique can be described as a finite difference method and involves measuring the change in Y while varying a particular X_i and holding all other input quantities constant. This is often used when there may not be a model available for a particular process but a rudimentary uncertainty analysis is required, or if the computational power is available to recompute the finite differences when the estimates change. Typically, the sum of the estimate and the standard uncertainty of each input quantity $\bar{x}_i + u(\bar{x}_i)$ is used, although a more rigorous version also includes the standard uncertainty subtracted from the estimate $\bar{x}_i + u(\bar{x}_i)$ to check for asymmetry. Because only two points are used to solve for each sensitivity coefficient (the estimate and the estimate plus standard uncertainty), this uncertainty propagation also linearises the measurement model.

If all of the input quantities are considered independent and the standard uncertainty was chosen as the value with which to perturb the input quantities, then by subtracting the estimate of the measurand from each sample and adding the results in quadrature, the combined standard uncertainty in the measurand can be obtained:

$$u_c(\bar{y}) = \sqrt{\sum_{i=1}^{N} ((\bar{x}_i + u(\bar{x}_i)) - \bar{x}_i)} \tag{3.18}$$

3.2.4 Expanded Uncertainty and Coverage Intervals

Although it is recommended to express a result with combined standard uncertainty $u_c(\bar{y})$, it is often required, especially in safety critical applications, for the uncertainty to encompass a larger fraction of the distribution of values that could reasonably be attributed to the measurand.

An expanded uncertainty, U, is instead used and is related to the combined uncertainty by $U = ku_c(\bar{y})$ [4, p. 6.2.1]. The multiplying factor, k, is termed the coverage factor and is typically in the range 2 to 3, often either of those two integer values and is defined by specifications or standards relating to the application. Using expanded uncertainties, the result can be expressed as $\bar{y} \pm U$, which is a popular format for datasheets and specifications.

To obtain a coverage factor that states a probability (e.g. 95%) that the true value of a measurand is within the associated interval is not straightforward, and depends on the probability distribution of the measurand. If all input quantities are Category A uncertainty components and the measurement model is linear, then the measurand distribution can be assumed to be Gaussian. In this case, the coverage interval is known as a confidence interval and can be given as a percentage by $\mathrm{erf}(z/\sqrt{2}) \times 100$, where $\mathrm{erf}(x)$ is the Gauss error function of x.

In situations where the above conditions cannot be met, a level of confidence can be obtained by calculating the effective degrees of freedom ν_{eff} of the distribution of the measurand. This process is explained in Annex G of [5]. For Monte Carlo propagations with sufficient samples, the confidence interval can be found by analysing the distribution of the measurand and obtaining the deviation from the estimated value which encompasses the desired percentage of samples (e.g. 95%).

3.3 Sensitivity Analysis

A benefit of propagating uncertainties through the measurement model is that an analysis of the sensitivity of the measurands to each input quantity can be performed. The sensitivity coefficients obtained from the measurement model can either be compared directly or multiplied by the standard uncertainty of the respective input quantity, in order to obtain an uncertainty figure for the measurand which can be compared with those calculated for other input quantities. This method is similar to the finite difference propagation technique described in 3.2.3.3, which can also be used to perform a sensitivity analysis. Because the input quantities are perturbed from their estimate sequentially (while all others are held at their estimate), this form of sensitivity analysis is termed “sequential perturbation”.

The results of the sensitivity analysis can be very useful to the metrologist. Not only can the relative impact of different input quantity uncertainties be reviewed, but also complicated behaviour in the combined standard uncertainty may be better understood; Figure 3.7 shows an example. Sensitivity analyses are also an efficient approach to reducing combined standard

uncertainty. Once input quantities with dominant contributions have been identified they can be targeted for improvement – or in some cases an alternative measurement model can be used which avoids them.

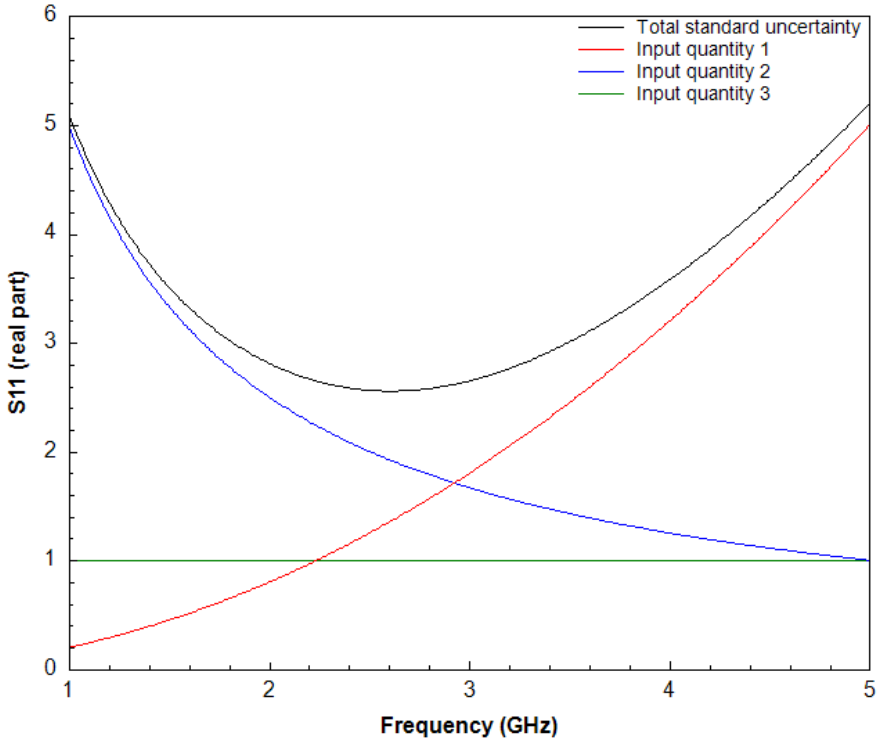


Figure 3.7: An example of results from a sensitivity analysis which reveal the origins of the complicated behaviour of the combined standard uncertainty with respect to a variable (in this case frequency).

3.4 Conclusions

This chapter presented how measurements underpin modern life, supporting trade and commerce and facilitating new discoveries in science and engineering. Through traceability and the unit system, the evaluation and management of uncertainty in measurements generates confidence and trust. In an attempt to standardise the definition and representation of measurement uncertainties, an internationally-used guidance document, the GUM, offers rigorous methods to evaluate these uncertainties. However, the GUM continues to be developed, and recently an inconsistency was created in the evaluation of Category A uncertainty components. This chapter

has reviewed the inconsistency from the objective of electromagnetic measurements, an area of metrology where the effects have been shown to be potentially significant.

Three methods for propagating uncertainty through a measurement model to determine the combined standard uncertainty of the measurands were described. Although the Monte Carlo method preserves the most information about both the uncertainties of the input quantities and the measurement model, the higher computational effort can be prohibitive in some cases. Instead, the LPU provides two linear alternatives, which are often much more efficient but require validation to ensure that the measurement model can be treated as linear.

The idea of expanded uncertainty and coverage intervals was introduced, these being met frequently in Category B uncertainty components defined from datasheets and specifications. A confidence interval is straightforward to calculate if the measurement involves only Category A uncertainty components and has a linear measurement model, or if a Monte Carlo propagation is used and the probability distribution of the measurand can be attributed to a standard type. In other cases, a coverage interval can be calculated using knowledge of the input quantities and further guidance from the GUM.

Finally, this chapter described sensitivity analysis, which can be carried out using results from the LPU procedure. The framework presented in this thesis utilises a sensitivity analysis to allow the user to examine and attempt to minimise significant sources of uncertainty, which is especially important in sensitive electromagnetic measurements such as those made on-wafer.

References

[1] UKAS. Our role, [Online]. Available: http://www.ukas.com/about/our-role/ (visited on 01/07/2019).

[2] BIPM. (Nov. 10, 2016). Draft of the ninth SI brochure, [Online]. Available: http://www.bipm.org/utils/common/pdf/si-brochure-draft-2016b.pdf (visited on 01/07/2019).

[3] International Organization for Standardization, *Guide to the expression of uncertainty in measurement*, 1st ed., Geneva, 1993.

[4] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML, *Guide to the Expression of Uncertainty in Measurement JCGM 100:2008 (GUM 1995 with minor corrections)*, 2008. [Online]. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM%20100%202008%20E.pdf.

[5] ——, *International Vocabulary of Metrology Basic and General Concepts and Associated Terms (VIM) JCGM 200:2012*, 2012. [Online]. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM%20200%202012.pdf.

[6] ——, *Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement" – Propagation of distributions using a Monte Carlo method JCGM 101:2008*, 2008. [Online]. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM%20101%202008%20E.pdf.

[7] ——, *Supplement 2 to the "Guide to the Expression of Uncertainty in Measurement" – Extension to any number of output quantities JCGM 102:2011*, 2011. [Online]. Available: http://www.bipm.org/utils/common/documents/jcgm/JCGM%20102%202011%20E.pdf.

[8] N. Metropolis and S. Ulam, "The monte carlo method," *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, Sep. 1949. DOI: 10.2307/2280232.

[9] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 236, no. 767, pp. 333–380, Aug. 1937. DOI: 10.1098/rsta.1937.0005.

[10] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. Chapman and Hall, 2013, 675 pp.

[11] D. R. White, "In pursuit of a fit-for-purpose uncertainty guide," *Metrologia*, vol. 53, no. 4, S107–S124, Jun. 2016. DOI: 10.1088/0026-1394/53/4/s107.

[12] R. Kacker, B. Toman, and D. Huang, "Comparison of ISO-GUM, draft GUM supplement 1 and Bayesian statistics using simple linear calibration," *Metrologia*, vol. 43, no. 4, S167–S177, Aug. 2006. DOI: 10.1088/0026-1394/43/4/s02.

[13] R. N. Kacker, "Bayesian alternative to the ISO-GUM's use of the Welch–Satterthwaite formula," *Metrologia*, vol. 43, no. 1, pp. 1–11, Nov. 2005. DOI: 10.1088/0026-1394/43/1/001.

[14] R. Kacker and A. Jones, "On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent," *Metrologia*, vol. 40, no. 5, pp. 235–248, Sep. 2003. DOI: 10.1088/0026-1394/40/5/305.

[15] W. Bich, "Revision of the 'guide to the expression of uncertainty in measurement'. why and how," *Metrologia*, vol. 51, no. 4, S155–S158, Jul. 2014. DOI: 10.1088/0026-1394/51/4/s155.

[16] L. Stant, P. Aaen, and N. Ridler, "Comparing methods for evaluating measurement uncertainty given in the JCGM 'evaluation of measurement data' documents," *Measurement*, vol. 94, pp. 847–851, Dec. 2016. DOI: 10.1016/j.measurement.2016.08.015.

[17] C. Elster, W. Wöger, and M. G. Cox, "Draft GUM supplement 1 and Bayesian analysis," *Metrologia*, vol. 44, no. 3, pp. L31–L32, May 2007. DOI: 10.1088/0026-1394/44/3/n03.

[18] K. Klauenberg and C. Elster, "The multivariate normal mean – sensitivity of its objective Bayesian estimates," *Metrologia*, vol. 49, no. 3, pp. 395–400, Apr. 2012. DOI: 10.1088/0026-1394/49/3/395.

[19] M. Cox and K. Shirono, "Informative Bayesian type a uncertainty evaluation, especially applicable to a small number of observations," *Metrologia*, vol. 54, no. 5, pp. 642–652, Aug. 2017. DOI: 10.1088/1681-7575/aa787f.

4 Evaluating Uncertainty in Vector Network Analyser Measurements

4.1 Introduction

The dramatic growth of radio-based devices and applications over the last 50 years has led to the VNA becoming a critical instrument in most laboratories. Many of these applications required both accurate and reliable measurements from these VNAs. This is particularly so in areas such as manufacturing, calibration and testing. This is often driven by requirements given in international Quality Management documents such as the ISO 9000 series of standards [1] (for manufacturing and process control) and the ISO 17025 standard [2] (for calibration and testing).

The requirements given in these international standards are for measurements that can be demonstrated as fit-for-purpose (in terms of the achievable level of accuracy, etc) and made traceable to the international system of units [3], [4]. These requirements were not trivial for a VNA due to the complicated nature of the VNA's operating principles, for example the calibration mathematics. Combined with the available computing power and cost at that time, a full or rigorous evaluation of uncertainty for VNA measurements, per for example the ISO GUM document, was difficult and time-consuming. This led to much work by experts in this field to develop easier methods that addressed these needs in ways that were suitable for use by end-users in the manufacturing, calibration and testing communities. Much of this work was undertaken by the ANAMET Technology Group (www.npl.co.uk/anamet) during the 1990s. This resulted in a series of reports [5]–[7] describing the development of a guidance document that gave a procedure for assessing the performance of calibrated VNAs. The resulting guidance document [8] was published by the European co-operation for Accreditation (EA, www.european-accreditation.org) so that laboratories operating to the ISO 17025 standard and/or ISO 9000 series of standards

could implement the method for their own purposes. Ownership of this EA document was later transferred to the European Association of National Metrology Institutes (EURAMET) and re-published [9] as part of their Calibration Guides series of documents. This document, along with the recent updated version, is available as a free download from the EURAMET web-site: www.euramet.org.

In addition to the EURAMET guide, VNA manufacturers have also produced their own advice for users to estimate the combined standard uncertainty in their measurements [10] and provided software tools in some cases [11], [12]. Often this advice is based on the same methods presented in the EURAMET guide.

In more recent years rigorous evaluations of VNA uncertainty have become possible, through both the efforts of NMIs and access to greater computing resources. The difference between the previous approaches (which we will call "residual error" evaluations) and rigorous evaluations concerns the way in which uncertainty contributed by the VNA calibration is estimated and included in the measurement model. The two methods will be discussed later in this Chapter. Figure 4.1 illustrates the general structure of VNA uncertainty sources and evaluation, from which each component will now be explained.

4.2 VNA Measurement Model Input Quantities

4.2.1 Calibration Standards

In order to perform the calibration, or error correction, of a VNA as described in Chapter 2, we must compare measurements of impedance standards to definitions of their true values in order to obtain error coefficients. It is interesting that although sources of error in the calibration cover both systematic and random types, when the calibration is performed and these quantities are measured, any random errors are captured in the evaluated combined uncertainty, which itself is purely systematic.

4.2.1.1 Definitions

Because the definitions of impedance standards are based on prior measurements, and all measurements include uncertainty, they are also included as a source of uncertainty in the VNA measurement model. We will now look at how uncertainty can be included in the three types of standard definition commonly used in VNA calibration routines.

Figure 4.1: Structure of VNA uncertainty evaluation. Input quantities are shown at the top of the diagram, grouped where applicable. Uncertainties from the calibration are either evaluated directly (rigorous/full evaluation) or are estimated after the calibration has been performed ("residual error" evaluation). Typically the calibration measurement model is processed as part of the VNA measurement model, but it is shown separately here to distinguish the difference between VNA uncertainty evaluation methods.

Databased Definitions The simplest definition of an impedance standard is the databased definition. The standard is characterised by measurement on a VNA which has been calibrated to a suitable accuracy, i.e. using standards which are at an appropriate position in the traceability chain (closer to the primary standard). The combined standard uncertainty of the characterisation is then provided with the estimates as the definition of the standard.

It is important to characterise the standard across a suitable frequency range, usually the entire range for which it can be applied. In addition to the combined standard uncertainty, the covariances between the real and imaginary components of the measured s-parameters can also be included in the standard definition for more accurate uncertainty evaluations in measurements using the standard [RidlerSalterEtc]. If the standard will be used to calibrate VNAs producing results to be used in time-domain studies or multi-harmonic studies (i.e. nonlinear work), then ideally covariances between s-parameters for each frequency should be included. However, this is something which would consume a lot of data storage and the saving of this information is not supported by VNAs at this time.

Polynomial Model Definitions In order to reduce the information required to accompany the standard, one-port models are often defined as coefficients of a polynomial fit between reflection coefficient and frequency. This characterisation is performed by the manufacturer and the coefficients are available in the user manual or specification sheet for the impedance standards. Testing is performed by the manufacturer to ensure that the supplied standards do not deviate from this characterisation. Similar to the databased definition, care must be taken to ensure the characterised polynomial fit is valid across a suitable frequency range.

An example of a short-circuit polynomial fit, as defined in [13], is

$$L_{\mathrm{S}} = L_0 + L_1 f + L_2 f^2 + L_3 f^3, \tag{4.1}$$

$$Z_{\mathrm{S}} = j2\pi f L_{\mathrm{S}}, \tag{4.2}$$

$$\Gamma_{\mathrm{S}} = \frac{Z_{\mathrm{S}} - Z_{\mathrm{r}}}{Z_{\mathrm{S}} + Z_{\mathrm{r}}}, \tag{4.3}$$

where L_{S} and Z_{S} are the respective inductance and impedance of the short-circuit, L_n are coefficients provided by the definition of the standard, Z_{r} is the reference impedance and Γ_{S} is the reflection coefficient of the short-circuit at frequency f. A similar model is used for open-circuits, with the inductive component replaced with a capacitive one.

Although no manufacturers are yet providing uncertainties with the polynomial coefficients

supplied in their standard definitions, some have included the ability for future incorporation in the data file structure. In addition to uncertainty originating from the measurement of the standard, there will also be uncertainties relating to the error of the polynomial fit. Because of this, the polynomial definition is typically the least accurate type of standard definition, however it is popular due to the portability of the small number of coefficients which can be defined for the entire set of manufactured models.

Physical Definitions Typically understood as the most accurate, physically defined standards use robust geometric models to calculate their impedances from dimensional measurements. Models are available in coaxial transmission line for short-circuits, open-circuits and arbitrary line lengths. Good examples of these models and their derivations can be found in [13], [14, Appendix C]. Matched loads are more difficult to accurately model and so they are typically included as a databased standard, even if the other standards are physically defined. In order to avoid this issue, the TRL calibration can be used, which requires standards that can all be physically defined. This combination of calibration model and standard definition is commonly used by NMIs when performing traceable characterisations of other standards (i.e. databased definitions).

To obtain uncertainties in the reflection coefficients of physically defined standards, which are the measurands which we would then like to use as input quantities to the VNA uncertainty evaluation, an additional uncertainty evaluation must be performed. This takes the estimates and uncertainties of the dimensional measurements of the standards and propagates them through the measurement model (the geometric model relating dimension and frequency to reflection coefficient) to obtain the reflection coefficients with uncertainties required.

Dimensional measurements of the standards are usually supplied by the manufacturer for TRL calibration kits, but like the other definition types can be supplanted by recent measurements performed by NMIs or calibration providers. Another benefit of physically defined impedance standards is that their definition is valid over their entire valid frequency range, removing the risk associated with databased and polynomial definitions.

Covariance information is not required to be supplied by the definition because the dimensional measurements should be uncorrelated (they are separately manufactured parts). During the uncertainty evaluation to obtain the reflection coefficients of the standards, covariances between both the complex impedance components and those at different frequencies are calculated. Therefore physical definitions are well suited to both portability, accuracy and use with time-

domain or nonlinear measurements.

4.2.1.2 Measurements

During the calibration procedure the impedance standards are measured. Although the VNA is not yet calibrated, it is still performing a measurement (with reference to an arbitrary impedance) and so all of the sources of uncertainty additional to those from the calibration will be included. Sources such as noise are therefore counted several times during a VNA measurement due to their inclusion in the prior calibration measurements. Because of this, it is important to include correlations where possible for these sources as they will have a greater impact on the combined standard uncertainty of the result.

4.2.2 Noise

Noise plays an important role in RF and microwave engineering. It is a random effect which occurs at a fundamental level and cannot be corrected for, both in communication systems and instrumentation. Wireless transmission systems lose a large amount of signal through propagation and receivers must be designed to handle signals with a considerable amount of noise. Likewise, VNA transmission measurements of devices with high isolation or reflection measurements of devices with low insertion loss also exhibit a low signal-to-noise ratio. To obtain accurate results, it is therefore important for VNA manufacturers to provide instruments with low noise sources and receivers. To ensure that measurements which are more susceptible to noise are accurate, it is also important to quantify the noise associated with the measurements so that it can be included as a contribution to their combined standard uncertainties.

The amount of electrical noise in VNA measurements is a sum of thermal noise and contributions from the VNA components. Thermal noise, caused by the random motion of free electrons in a conducting material by heat, is specified as 4×10^{21} W/Hz (-174 dBm) [10]. The intermediate frequency bandwidth (IFBW) setting on the VNA controls the frequency range of signals being measured at each discrete point in the sweep, and can typically be set from 1 Hz to more than 10 kHz. It is this figure which multiplies the thermal noise to provide a theoretical minimum noise floor if we assumed the VNA did not contribute any noise itself. The disadvantage of a very low IFBW is that the VNA takes longer to perform the measurement, so there is a compromise between speed and accuracy. This provides a good example of where uncertainty evaluation can have a direct perceivable benefit to measurement accuracy - if the

IFBW is reduced the uncertainty due to noise should also reduce. By quantifying the uncertainty the engineer can make an informed decision as to what value to set the IFBW, which could have expensive time implications in a test environment.

In addition to the thermal noise, the VNA contributes to the noise level from the source, receiver and other test set components (i.e. local oscillator phase noise). Setting attenuators to lower values will improve noise as resistors add additional thermal noise to the measurement. The noise figure (NF) of a VNA is typically quoted in the specifications, and can be used with the thermal noise to calculate the theoretical noise level of the VNA measurement [10]:

$$L_{\mathrm{N}} = -174\mathrm{dBm} + NF + 10\log(S_{\mathrm{F}})\mathrm{dB} + 10\log\left(\frac{B_{\mathrm{IF}}}{\mathrm{Hz}}\right)\mathrm{dB} \tag{4.4}$$

For the purposes of uncertainty evaluation, VNA noise level is measured on the instrument itself and is typically included in the measurement model as two input quantities, noise floor and trace noise. It is important that the VNA settings (IFBW, test port power, averaging) are the same as will be used for the DUT measurements, and that no calibration is applied.

4.2.2.1 Noise Floor

The noise floor is a term used to describe the noise present in the measurement with no external signal present. It can be measured with a matched load connected directly to the test port and should not include any noise contribution from the VNA source. Alternatively it can be measured from the transmission measurements of a 2-port VNA while short-circuits or open-circuits are connected for the trace noise measurement described below. Many measurements are made ([9] suggests a few hundred) and the standard deviation is calculated for each frequency point.

4.2.2.2 Trace Noise

Trace noise includes noise contributions from the VNA source and is a function of the measured power at the receiver. It is measured with a short-circuit or open-circuit connected directly to the test port.

4.2.3 Repeatability

Both the DUT connections and movement of the VNA port extension cables contribute uncertainty to the repeatability of VNA measurements. These errors vary when a different device is

connected to the test setup, and are included for both the DUT measurement and the measurements of the impedance standards during calibration.

4.2.3.1 Connections

Connector repeatability is a fundamental quality of precision RF and microwave connectors. It can have a significant impact on DUT measurements if the response of the connector varies between multiple connections, and this perturbs the reference impedance of the calibrated VNA away from 50 Ohms if this occurs between connecting calibration standards.

In order to measure connector repeatability uncertainty, a short-circuit can be connected to a calibrated VNA and the reflection coefficient measured. Measurements are then repeated, reconnecting the short at a different azimuthal rotation each time. This should be done at least three times with rotations of 120 deg, although some guidance recommends up to 16 times[9].

4.2.3.2 Cable Stability

When test port cables are flexed, their physical dimensions are perturbed, which in turn affects their s-parameter response. Specialist VNA test port extension cables are provided to mitigate the effects of cable flexure, although they do not remove them. It is therefore advisable to restrict the movement of these cables using clamps and supports, and also to wait a suitable time (typically 30 seconds or more) for stresses in the internal dielectric to settle.

Cable stability can be characterised in different ways depending on the VNA measurement model used. The general method is to perform repeated measurements while moving the cable between a range of positions which cover the maximum extent the cable will move during measurements of the calibration standards and DUTs. Specifically, some methods connect a short-circuit to each test port cable, and others connect the cables together to perform the characterisation (if used for a two-port measurement).

4.2.4 Drift

Drift does not strictly fit into the random or systematic error category, because it can vary over the course of measurements but is not random in nature. The dominant cause of drift is environmental conditions, especially temperature variations. Temperature affects the physical dimensions and s-parameters of RF components in the test setup and also the behaviour of the electronic devices inside the source and receiver - both of which perturb the result of the

measurement.

Drift is an important source of uncertainty to include for measurements occurring over a long period of time after calibration, especially those in a test environment or in standards labs while performing a large number of repeat measurements. In order to minimise the effects of drift, the room which the VNA is located in should have good temperature control and all the equipment should have been powered on for at least 24 hours for their internal temperatures to settle.

In order to measure uncertainty due to drift, many repeat measurements are performed over a long period (24 hours or more). This procedure can be automated so that an operator is not required for the entire period. An alternative method of including uncertainty due to drift can be achieved by performing calibration measurements of standards both before and after the DUT measurements. The calibrations can be averaged and a crude standard deviation obtained.

4.2.4.1 VNA Linearity

One source of VNA measurement uncertainty which is still included in some evaluations is that from the nonlinearity of the receivers. However, in modern VNAs automatic level control corrections are sufficiently accurate to allow this error to be neglected [15], [16].

4.3 Simplified Residual Model for VNA Uncertainty Evaluation

4.3.1 Method

The EURAMET Guide [9] presents a process for evaluating the uncertainty of measurements performed on a calibrated VNA, allowing users to verify that values measured using the instrument are of acceptable accuracy. This process involves measuring a selection of dominant contributions to measurement uncertainty and combining them appropriately. Contributions include both systematic errors, which remain constant over the period of measurements, and random errors, which do not. The error model for voltage reflection coefficient (Γ) measurements performed with a VNA is represented in [9] by the following equations for one-port (1) and two-port (2) measurements:

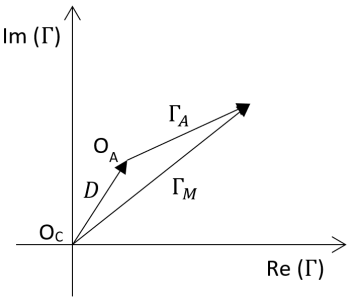
$$U_\Gamma = D + T\Gamma + M\Gamma^2 + R_\Gamma \tag{4.5}$$

$$U_\Gamma = D + T\Gamma + M\Gamma^2 + R_\Gamma + S_{21}^2\Gamma_L \tag{4.6}$$

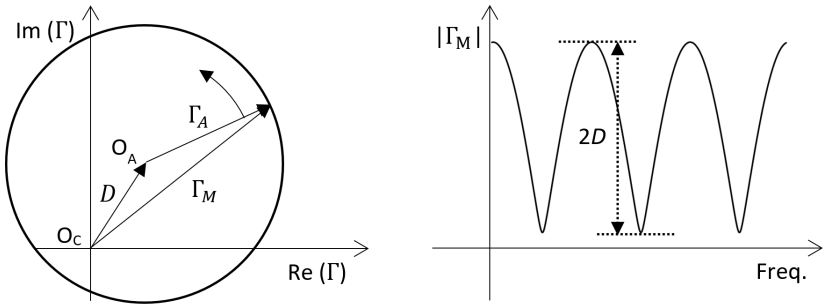
where U_Γ is the combined uncertainty in the measurement, D is the residual directivity, T represents the effect of tracking and nonlinearity, M is the residual test port match (TPM), Γ is the measured voltage reflection coefficient, R_Γ is the sum of all the random contributions, S_{21} is the nominal attenuation of the device-under-test (DUT), and Γ_L is the residual load match. The most significant systematic error contributors to the measurement uncertainty are, in most cases, the directivity and TPM.

To measure Γ, the VNA must separate reflected and incident voltage waves and then sample them using complex receivers. However, various components in the signal path may cause a portion of the incident wave to leak into the reflected wave receiver without having reached the DUT. This directivity error should be removed by applying correction terms extracted during the VNA calibration. However, as no calibration will be perfect, some residual directivity error will remain (referred to as effective directivity in [9]). To measure the residual directivity, a matched load can be connected to the test port being assessed. This should theoretically reflect none of the incident wave and the only voltage present at the reflected wave receiver should be due to the residual directivity. In practise, the match of the load will never be perfect, so it is likely that using this method the residual directivity will typically be either over- or underestimated. An improved method, used in [9] and widely accepted for use with coaxial measurements, is called the 'ripple extraction technique'. This uses a similar principle to measure the residual directivity, but significantly improves the accuracy of the residual error evaluation. An illustration of its method is provided in Figure 4.2.

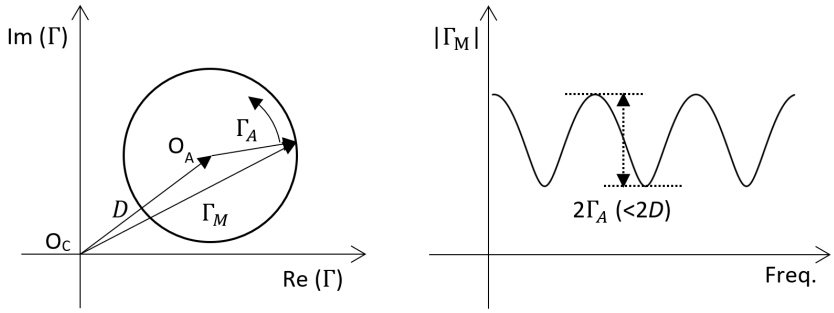
To perform the ripple extraction technique, a short length of line is connected to the test port, to the end of which is added a near-matched load covering the frequency range under test. The critical dimensions of the line section (length and radii) should be traceable to national standards and have a characteristic impedance identical to that of the VNA setup. For these reasons a beadless airline is suggested in [9]. The load can be either the same as that used for calibration or another with $0.1 \geq |\Gamma| \geq 0.2$ (in linear units) to ensure that $|\Gamma| \geq |D|$. If $|\Gamma| < |D|$, then the measured residual directivity will be underestimated as explained by Figure 4.2. If the calibration matched load is used for the measurement, the small reflection from a second connection and any loss in the airline will cause Γ to be greater than the residual directivity



(a) When measured on the calibrated VNA, a perfect matched load would reveal the actual origin (O_A) on a polar plot of Γ as offset from the calibrated origin (O_C) by the residual directivity (D). If a realistic matched load offset by a line section is instead measured, Γ as measured by the VNA (Γ_M) will be the sum of the residual directivity D and the actual Γ (Γ_A).



(b) As Γ_M is measured across a swept frequency range, the phase change in the line increases causing the phase of Γ_A to sweep also. This rotates Γ_A, resulting in ripples in the plot of $|\Gamma_\mathrm{M}|$ against frequency. The magnitude of the ripples is equal to $2D$.



(c) However, if $\Gamma_\mathrm{A} < D$, then the ripple magnitude is now $2\Gamma_\mathrm{A}$ instead of $2D$ and the residual directivity as evaluated using the ripple extraction technique would be underestimated.

Figure 4.2: A graphical representation of the ripple extraction technique, including a possible failure mechanism.

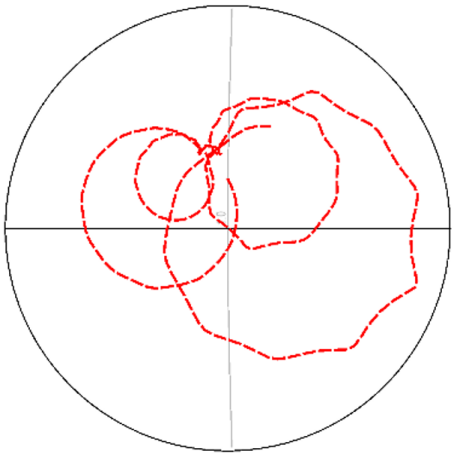
from the original measurement of the load. Alternatively, because $|\Gamma| < 0.1$ for the matched load used for calibration, using another load with a known higher $|\Gamma|$ ensures that there is no underestimate. Once the instrument has been configured, Γ is measured and the magnitude plotted against frequency using a linear scale. A ripple will then be visible on the trace, from which the residual directivity can be calculated from:

$$D = \frac{\mathrm{MRA}_{\text{matched-load}}}{2} \tag{4.7}$$

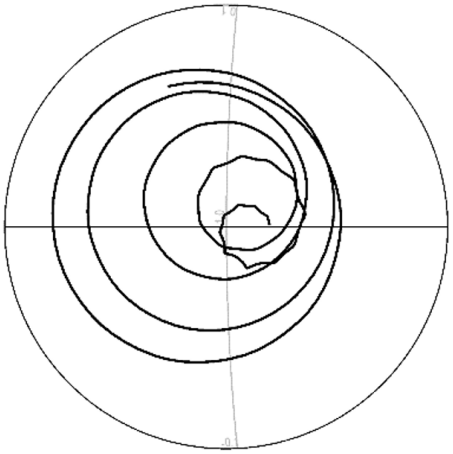
For coaxial measurements as specified in [9], there is a high probability that the condition required to avoid underestimation of $|D|$ is met. However, in order to assess the suitability of the technique in waveguide a method of assessing this condition has been used. By examining either a complex plot (polar or Smith chart) or a phase plot, the geometric symptom shown in Figure 4.2 can be identified. When using a complex plot, the origin should lie within the circumference of the reflection coefficient trace for a valid determination of the residual error to be achieved. For any frequency range where it does not, the ripple technique provides an underestimation of the residual error. When using a phase plot, there will be regular wrapping of the reflection coefficient phase for frequency ranges where the residual error is correctly measured, whereas when underestimation occurs the phase will vary by $< 180°$ per period. An example of both plots are shown in Figure 4.3, where the result from the different load indicates an accurate residual directivity estimate across the entire measured spectrum, but the result from the calibration load shows an underestimate of the residual directivity is likely between 16 and 22 GHz. Either of these methods can be used to identify when a calibration and the ripple extraction technique needs to be repeated. If the repeat measurements still fail the test, then the choice of loads may need to be altered.

TPM is caused by imperfections in the impedance match between components in the VNA setup. This causes delayed reflections that interfere with the DUT measurement and can provide false values. Calibration also corrects for TPM, but as with directivity some residual error will remain. To measure residual TPM, a short circuit is connected to the test port being assessed. This should reflect the entire incident signal and therefore maximise any reflections in the VNA setup. If residual TPM error is present then the measured Γ will be < 1. However, the short circuit may not provide a perfect reflection and so the ripple extraction technique is favoured for this measurement also.

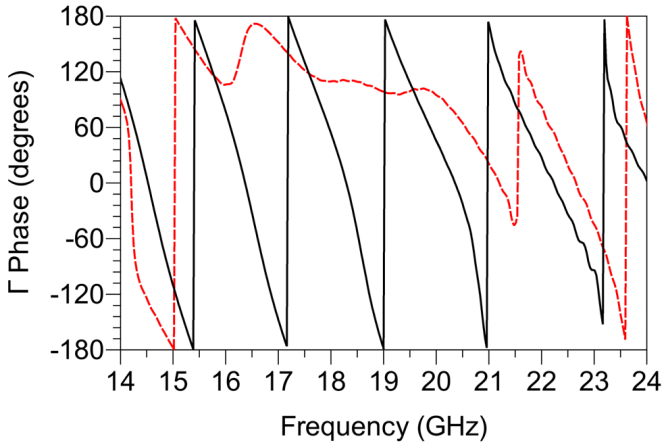
To measure residual TPM using the ripple extraction technique, the same procedure as for residual directivity is followed but the matched load at the end of the line is replaced by a short



(a) A Smith chart representation, magnified about the origin, of the calibration load result.

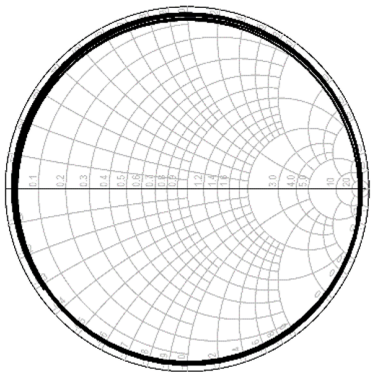


(b) Similar Smith chart representation of the different load result. The intersection of the two grid lines on the Smith charts represents the origin of the plot ($|\Gamma| = 0$), and both are scaled separately for clarity.

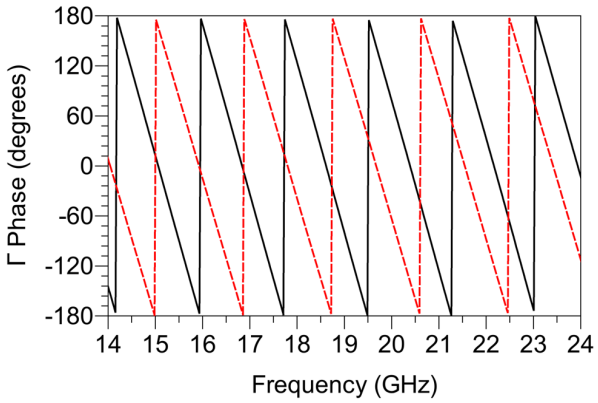


(c) Phase plot of both results, with the calibration load represented by the red dotted trace and the different load by the black solid trace.

Figure 4.3: Measurements of residual directivity plotted over the frequency range of 14–24 GHz performed in coaxial transmission line using both the load used for calibration and a load from a different calibration kit.



(a) Smith chart representation of the result when using the calibration short-circuit. The result when using a short-circuit from a different calibration kit appears almost identical.



(b) Phase plot of the results when using the calibration short-circuit (dotted red trace) and the different short-circuit (solid black trace).

Figure 4.4: Measurements of residual TPM plotted over the frequency range of 14–24 GHz performed in coaxial transmission line.

circuit. A similar plot is acquired and the residual TPM is given by:

$$M = \frac{\mathrm{MRA}_{\text{short-circuit}}}{2} \tag{4.8}$$

Since the reflection coefficient for this measurement should be close to 1, there is no risk that this value will be greater than the true residual TPM and cause an underestimate as can be the case for residual directivity. This is shown in Figure 4.4, where the origin of the Smith chart is clearly inside the circular trace (Figure 4.4a) and the phase consistently wraps across the entire measured spectrum (Figure 4.4b).

To perform the technique for both described residual error sources requires just three components: A short circuit, matched load, and a short section of line. These components are realizable in both coaxial and rectangular waveguide, so the technique should be physically possible to perform in waveguide setups. The technique can be applied to any type of calibration – for example three-known-loads and thru-reflect-line (TRL). In coaxial, the short-open-load-thru (SOLT) variant of the former is used. However, in waveguide an open circuit is not straightforward to realise or widely adopted, so a common variant of SOLT calibration which uses an offset short (SOSLT) will be used instead.

4.3.2 Application to Waveguide VNAs

Although both the EURAMET guide and VNAs using rectangular metal waveguide test ports have both existed for many years, there was no published evidence that the methods employed in the guide (the ripple technique) could be successfully applied to those measurements. The author undertook an investigation into this during the degree where they compared residual error measurements in coaxial line to those in waveguides at frequencies ranging from 8.2 GHz to 750 GHz (submillimetre-wave). This section presents the results of the subsequent papers [17], [18].

4.3.2.1 Coaxial Transmission Line and Microwave Waveguide

The ripple extraction technique was first performed on coaxial line in accordance with the EURAMET Guide [9] instruction as described in the previous section. The guide itself provides a range of typical values for both residual directivity and TPM ripple measurements with which our results can be compared to ensure that the process was followed correctly. All measurements presented in this investigation were acquired using a Keysight 5247A PNA-X instrument fitted with 1.85 mm ports attached to flexible port extender cables with rugged connectors. The coaxial measurement setup used a 75 mm beadless airline and the calibration kit matched load and short circuit. Figure 4.5 shows the ripple trace obtained by plotting $|\Gamma|$ against frequency for residual directivity and TPM measurements using an SOLT calibration. Apart from the dominant ripple, other variations in $|\Gamma|$ are due to the imperfect response of the matched load and the beadless line. The results of the measurements, along with the expected ranges provided in [9], are presented in Table 4.1. It can be seen that the measured values for the coaxial line setup fall within the typical ranges specified by [9].

The same method was applied to two types of centimetre band waveguide, WR-90 and WR-42. These waveguides have usable frequency ranges of 8.2–12.4 GHz and 18.0–26.5 GHz respectively. To avoid the effects of non-propagating (evanescent) modes created by the waveguide to coaxial adapter, an appropriate length of straight waveguide was attached to each adapter where possible and the measurement planes defined at the end of the lines. The results of the ripple extraction technique for the two waveguide sizes are shown in Table 4.2:

4.3.2.2 Millimetre-wave Waveguides

To perform measurements at frequencies above 50 GHz, a range of external frequency extender heads were attached to the VNA. These extender heads included a line section attached to each

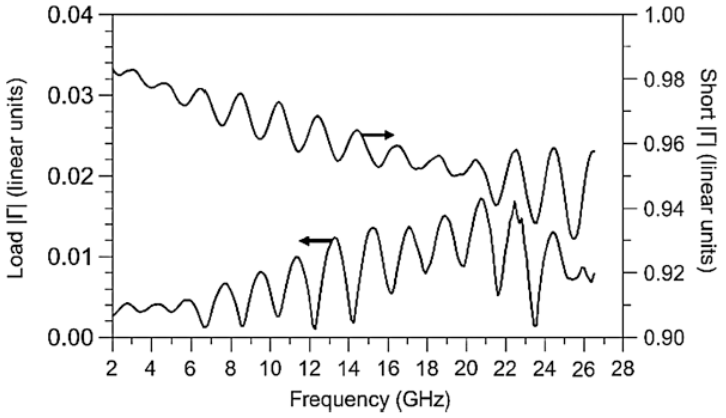


Figure 4.5: Magnitude of the reflection measurement of a matched load and a short-circuit offset by a length of coaxial line. The VNA was calibrated using the SOLT method.

Cal. type	Residual directivity		Residual TPM	
	Port 1	Port 2	Port 1	Port 2
SOLT	0.008	0.009	0.014	0.010
TRL	0.008	0.007	0.002	0.002
EURAMET guide [9]	0.002–0.02	0.002–0.02	0.005–0.02	0.005–0.02

Table 4.1: Residual directivity and TPM values obtained for 3.5 mm coaxial line VNA calibrations as measured by the ripple extraction method. Both SOLT and TRL calibration techniques were assessed. The range of representative residual error values from [9] has also been included for comparison.

Frequency, GHz	Waveguide size	Cal. type	Residual Directivity		Residual TPM	
			Port 1	Port 2	Port 1	Port 2
8.2–12.4	WR-90	SOSLT	0.005	0.004	0.007	0.006
		TRL	0.003	0.006	0.002	0.002
18–26.5	WR-42	SOSLT	0.003	0.004	0.010	0.005
		TRL	0.002	0.002	0.002	0.001

Table 4.2: Residual directivity and TPM values of WR-15 and WR-05 waveguide VNA calibrations as measured by the ripple extraction method. Both SOSLT and TRL techniques were used to calibrate the VNA.

Frequency, GHz	Waveguide size	Cal. type	Residual directivity		Residual TPM	
			Port 1	Port 2	Port 1	Port 2
50–75	WR-15	SOSLT	0.002	0.002	0.018	0.017
		TRL	0.002	0.002	0.009	0.007
140–220	WR-05	SOSLT	0.008	0.009	0.019	0.024
		TRL	0.007	0.008	0.021	0.015

Table 4.3: Residual directivity and TPM values of WG25 and WG30 waveguide VNA calibrations as measured by the ripple extraction method. Both SOSLT and TRL calibration techniques were measured.

test port of suitable length to avoid effects caused by evanescent modes that may exist close to the test port. To study the performance of the ripple extraction technique at millimetre wavelengths, WR-15 and WR-05 waveguides were chosen. These waveguides have operating frequency ranges of 50–75 GHz and 140–220 GHz respectively. The results of the ripple extraction are shown in Table 4.3:

4.3.2.3 Submillimetre-wave Waveguides

The final stage of the investigation studied the ripple extraction technique when applied to submillimetre wavelength VNA setups. The waveguide chosen for these measurements was in the 500–750 GHz band (WR-1.5) for which only one frequency extender head was available. Owing to the requirement for a through standard when using the TRL calibration method, only a three-known-loads calibration was performed, which was the one-port version of SOSLT (SOSL). The line section used was ∼25.4 mm in length and was part of a calibration and verification kit manufactured by Virginia Diodes, Inc. All waveguide flanges, apart from the short-circuit, used precision alignment dowels located above and below the aperture. Figure 4.6 shows the setup used during the measurement of residual directivity, with a line and match connected to the frequency extender head.

It can be seen that the residual directivity is significantly smaller than the residual TPM. By assessing the phase wrapping of the ripple trace obtained for evaluating residual directivity, as shown in Figure 4.7, the ripple extraction technique appears to be subject to the failure mechanism described in the previous Section 4.3.1 and illustrated in Figure 4.2c. The lack of phase wrapping across the entire operating bandwidth shows that the ripple extraction technique

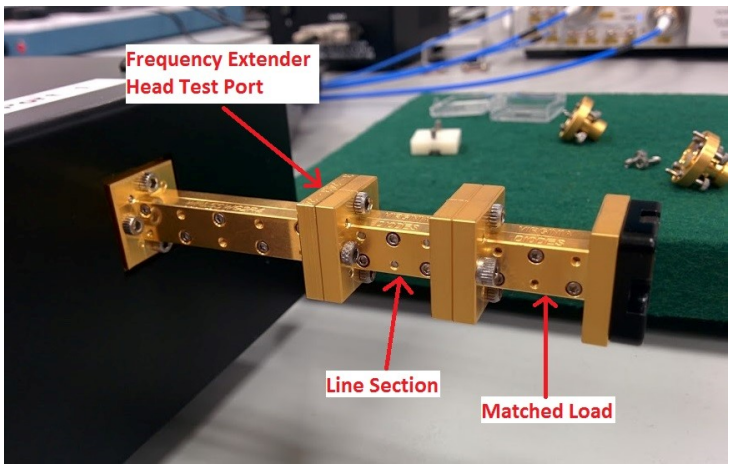


Figure 4.6: Test port setup used for residual directivity measurements in WR-1.5 waveguide.

was not operating within the required assumptions necessary for the technique to be valid, and therefore provided an underestimate of the residual directivity. The calibration kit used for this experiment included two full sets of standards, so the matched load was swapped and the ripple extraction technique was repeated. The issue was not resolved by this change, so the calibration was repeated, this time using the other matched load. The ripple extraction technique was then performed with the matched load used for the original calibration. However, no combination of these components provided a valid residual directivity value as assessed by the phase wrapping method. A likely cause of this effect is the poor connection repeatability inherent in this waveguide size, using typical precision UG-387 flanges. If the waveguide apertures have a greater misalignment during calibration than when the ripple extraction technique is performed, the effect of the discontinuities can cause the calibration matched load to appear to have a higher $|\Gamma|$ than the one used for the ripple extraction technique (even if the opposite were in fact true). These conditions cause the residual directivity to be underestimated as explained in Section 4.3.1.

4.3.2.4 Waveguide Discontinuities

When a discontinuity is present between two sections of rectangular waveguide, an increased reflection will be seen at the location of the join. There are three types of discontinuity possible in rectangular waveguide: E-plane and H-plane lateral displacements, angular displacement, and corner rounding. A report produced by Bannister *et al.* [19] presented the effects of these

Frequency, GHz	Waveguide size	Cal. type	Residual directivity	Residual TPM
500–750	WR-1.5	SOSLT Cal. 1	0.021	0.142
		SOSLT Cal. 2	0.025	0.065

Table 4.4: Residual directivity and TPM values of two WR-1.5 waveguide VNA calibrations as measured by the ripple extraction method. Two similar types of calibration were performed using different standards from the same kit.

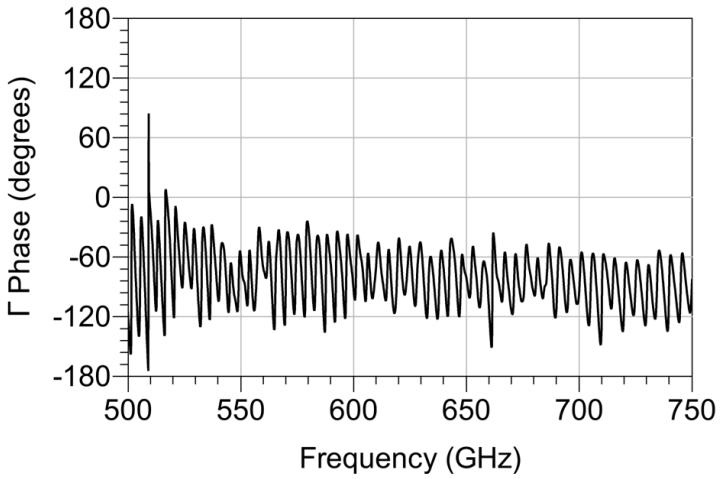


Figure 4.7: Phase plot for a residual directivity evaluation performed in WR-1.5 waveguide using an SOSLT calibration.

discontinuities at centimetre and millimetre wavelengths. Subsequent work by Kerr extended this using simulations [20], [21].

The error in Γ contributed to by the effect of such discontinuities is proportional to wavelength, and therefore also the aperture size of the waveguide. At submillimetre wavelengths, the error has been shown to be considerable [22], [23]. Recently, efforts have been made to improve the connection repeatability for waveguide at these wavelengths, and a new IEEE draft standard [24] presents three new connector types which significantly improves the alignment. Another potential improvement attempts to reduce misalignment errors during calibration by replacing the waveguide offset short with a radiating open, and new calibration algorithms have been developed to accompany this [25]. The open standard must be very well characterised, and the technique has not yet seen widespread uptake.

4.3.2.5 Discussion

In centimetre-wave waveguides (WR-90, WR-42), the ripple extraction technique provided values of residual directivity and TPM similar to those obtained in coaxial line. This is not surprising as providing a good quality match is used for calibration and the connections are well made and aligned, both errors should not have significant contributions outside of the VNA itself. When extended to millimetre-wave waveguide, the residual errors were found to be larger, but still within the recommended values for coaxial lines. At submillimetre wavelengths, the two calibrations that were performed resulted in significant differences in the residual errors. A likely cause for these differences is the effect of discontinuities in the waveguide components (due to poor connection repeatability and demanding mechanical tolerances) used during the calibration, especially the matched load and offset short. Additionally, by studying the phase of the measured Γ, it was shown that the ripple extraction technique was not operating within required assumptions necessary for the technique to be valid and was therefore underestimating the true value of the residual directivity. The cause of this may also be related to the poor repeatability of the waveguide connection, causing the Γ of the matched load used for the ripple extraction to be lower than that used for the calibration. Some effort were made to resolve this issue, but were unsuccessful. New improvements to submillimetre-wave waveguide flanges could significantly reduce this problem and allow the ripple technique to work more consistently with these very small waveguides.

A useful assessment method to test the validity of the ripple extraction technique was presented earlier in this paper, and should be performed whenever the technique is used. This

assessment method views Γ measured during the ripple extraction technique on either a phase plot, a polar plot or a Smith chart. When using a phase plot, this assessment has passed if the phase is seen to be wrapping across the operating bandwidth. When using a polar plot or Smith chart, this assessment has passed if the origin of the chart lies within the circumference of the trace. If the technique is deemed to have failed based on this assessment, the VNA should be recalibrated and the ripple extraction technique should be repeated. If this does not affect the assessment result, then the near-matched load or short-circuit used during the technique should be swapped with another (preferably known to have a higher Γ) and the ripple extraction technique repeated again. Only when this assessment has passed can the results from the ripple extraction technique be considered reliable.

4.3.2.6 Conclusion

This paper investigated the effectiveness of the ripple extraction technique when applied to rectangular waveguide measurements at centimetre, millimetre and submillimetre wavelengths. Typical values of residual directivity and TPM in these three ranges have been provided. For centimetre- and millimetre-wave waveguides the ripple extraction technique works as expected. Since the standards used in this investigation are commercially available and the calibration techniques are available on nearly all VNAs, the values of residual error given in this paper are considered representative and suitable for use as reference values by other users of similar types of measurement system, to compare their own values against. This is because the performance of these systems is usually dominated by the quality of the standards and the type of calibration technique that is used. However, this paper has also shown that the ripple extraction technique may not currently be a reliable way of measuring residual error in submillimetre wavelength systems. The effect of discontinuities at submillimetre-wave waveguide interconnections has been considered as a cause of this issue.

4.4 Rigorous Models for VNA Uncertainty Evaluation

4.4.1 Method

A rigorous evaluation of VNA uncertainty can be applied by using a measurement model which includes explicit relationships of the measurands to all input quantities. This is opposed to the residual model where the combined uncertainty from a collection of input quantities (relating to

the calibration) is approximated. By including explicit relationships for all input quantities, the uncertainty evaluation can be directly traceable to primary standards.

There are several formulations of rigorous VNA measurement models in use, and most of them have been written into a software framework due to their complexity. The differences between the models are due to optimising the same mathematical relationships for different methods of processing. This section will now describe three popular VNA uncertainty frameworks.

4.4.2 Keysight PNA-X Dynamic S-Parameter Uncertainty

The Keysight PNA-X Dynamic S-Parameter Uncertainty option is an extension to the software provided on Keysight PNA-X VNAs. This software extension uses a separate graphical window to allow the user to manage sources of uncertainty (input quantities), which are then referenced by the VNA software to evaluate the combined standard uncertainty in the s-parameter measurements it produces.

Uncertainties for input quantities are provided in the form of calibration kits, cables and test ports. Mechanical standards in calibration kits can be characterised with uncertainty information by either measuring them on another PNA-X with the uncertainty software option, or supplied with that information from the manufacturer if they are Keysight products. Likewise, electronic calibration units from Keysight also include uncertainty characterisations in their definition files and can be loaded directly into the PNA-X software. Cable stability can be measured in a guided process through the uncertainty software and assigned to that cable identity for future use. Likewise, the test-port is used as an identity to store measurements of noise floor and trace noise, which are again measured through a guided process.

A calibration using uncertainty information is then performed, after which the VNA will perform measurements as normal. However, with the calibration applied it is possible to also view uncertainty information in real-time (i.e. calculated after every sweep) on the display. This can be in the form of error bars and lines showing standard deviations for scalar measurands like magnitude and phase, and also uncertainty ellipses showing standard deviations of complex values on polar plots or Smith charts.

The measurement model used within the PNA-X Dynamic Uncertainty Option is not fully disclosed, but it is based on a multi-port formulation developed by Garelli and Ferrero [26]. This model rearranges the traditional calibration mathematics and creates new parameters which are very efficient for solving n-port calibration problems. This is a desirable property for commercial applications of VNAs where some communications products may require more than 24 ports to

fully characterise their s-parameters.

Uncertainty information is stored in several file formats. The ".u*n*p" format, similar to the ".s*n*p" format, contains s-parameter data for an n-port device along with uncertainties in their respective scalar components (either real/imaginary or magnitude/phase). The Databased Standard Definition ".dsd" format contains characterisations of calibration standards, which in addition to uncertainty data from the ".u*n*p" format includes information about the type of calibration standard which is used by the software to ensure calibrations are performed correctly. Finally, the ".sdatcv" file format can also be read from and written to, which is native to the VNA Tools II software explained shortly.

A more detailed overview of the methods used in the software can be found in the product Application Note [27].

4.4.3 METAS VNA Tools II

VNA Tools II is a piece of standalone software developed by the Swiss NMI, the Federal Institute of Metrology (METAS) [28]. Based on an uncertainty propagation library called UncLib [29], also by METAS, the software is not attached to a particular VNA but can instead drive instruments from several popular manufacturers, or post-process existing measurements.

The underlying uncertainty propagation library, UncLib, specialises in linear propagation of uncertainty according the method described in the GUM document introduced in Chapter 3. Monte Carlo propagation is also included in the library, but is not as well supported. VNA Tools II defines a rigorous VNA measurement model using mathematics functions from UncLib, to which it provides measurement data and processes the uncertainty evaluation [30].

In addition to the uncertainty evaluation, VNA Tools II allows users to store collections of characterised cables and calibration standards, as well as record all measurements in a journal for both short-term reference and archiving evidence of traceability.

To store uncertainty information for s-parameter measurements, a new file format was invented for VNA Tools II, which uses the ".sdatcv" filename extension. This file format stores s-parameter data with covariances between the real and imaginary components for each parameter at a particular frequency. Covariance data of parameters between frequency points is not saved as this can create very large files for typical VNA frequency sweeps, and is not useful when evaluating measurement uncertainty in the frequency domain for devices operating linearly. As mentioned earlier, this file format is also supported by the Keysight PNA-X Dynamic S-Parameter Uncertainty option, allowing standards to be used between the two different

software frameworks.

4.4.4 NIST Microwave Uncertainty Framework

The Microwave Uncertainty Framework (MUF) published by the US National Institute of Standards and Technology (NIST) is another software framework for evaluating uncertainty from an NMI [31]–[33]. Similar to VNA Tools II, the software does not need to be used with a specific brand of VNA because it can read existing measurement files (e.g. s-parameters) and perform calibrations during post-processing. The facility is available, however, to control Keysight VNA instruments remotely which can automate long measurement tasks for metrology studies.

Because the MUF was written to support general RF and microwave metrology research at NIST, many measurement models are included in the framework, of which VNA measurement models are just a few. The framework is provided as a collection of individual applications: **Post-Processor** contains a library of measurement models which can be evaluated using either Monte Carlo or finite difference (including sensitivity analysis) propagation techniques. **Model** allows two-port components to be defined, which can be cascaded together using **Cascade** and have their numerical parameters (e.g. length, resistance) assigned with statistical distributions using **Parameter**. One-port devices such as impedance standards are modelled as two one-port devices back-to-back. **Combine** takes a series of repeated measurements as an input and produces an estimate with associated uncertainty, according to classical probability methods. Finally, the **VNA Uncertainty Calculator** is a customised version of **Post-Processor** which includes a graphical layout to simplify loading of the many input quantities required for the VNA measurement models. Separate models are included for VNAs with up to eight ports, and six calibration types are available [34].

The MUF does not propagate uncertainty using covariance information stored in conventional matrices. Instead, at first chance it converts uncertainty information into a set of Monte Carlo samples. The order of these samples is carefully preserved throughout subsequent calculations, which has the effect of propagating covariance information throughout the uncertainty evaluation. For example, when a physically-defined impedance standard is modelled with parameters assigned statistical distributions, the MUF will sample from each distribution, evaluate the model and produce a sample to be used as an observation of an input quantity. This sample-based preservation of covariance information is very useful for input quantities with strong correlations - for example the real and imaginary components of a complex s-parameter measurement. Rather than statistically averaging a set of repeat measurements to assign a single covariance

value, as outlined in the GUM document, the repeat measurements are preserved so that the distribution of covariance information is also used during the uncertainty evaluation. For most types of uncertainty source in the MUF this benefit is not realised because Monte Carlo samples are synthesised from statistical definitions (e.g. models, parameters), but for large numbers of repeat samples the functionality is available. The side-effect of this benefit is that keeping all of the measurement data, or synthesised Monte Carlo samples, can consume a lot of memory. However, modern computing resources mean that the space requirement is no longer a difficulty.

One of the most flexible features of the MUF is that any user can define their own models and post-processors (both are measurement models). This allows the MUF to provide uncertainty evaluations for new methods of electromagnetic wave metrology and communication systems research [35], [36].

4.5 NVNA Uncertainty Evaluation

It was explained in Chapter 2 that in order to measure nonlinear devices using a VNA, the absolute power and phase of the waves at all test ports must be known. For a VNA to obtain this information, it must undergo additional calibration steps involving a phase reference and power meter. In addition, for mixer-based VNAs (the type used throughout this work), a stable phase reference must be connected to another port of the VNA and measured alongside any DUT measurements. Each of these additional requirements add input quantities to the NVNA measurement model when compared to that of the VNA, and these will now be discussed.

4.5.1 Phase Reference

The phase reference used during calibration contributes both systematic and random error sources. Because it is used as a calibration standard, systematic error due to inaccuracies of the characterisation is a primary concern. This is especially true for the phase calibration where the coefficients calculated from the measurements of the reference it's characterisation have a linear relationship with the calibrated DUT waves. This is different to the relative calibration of the VNA, where at least some of the coefficients are the result of information from several different impedance standards, therefore reducing the impact of a single inaccuracy in the characterisation or measurement of a single standard. In addition, for measurements of nonlinear microwave (and millimetre-wave) devices, harmonics of interest can be well above 10 GHz, where the wavelength in a transmission line is short and connector and cable repeatability errors (i.e.

random errors) can have a large effect on the accuracy of the phase calibration.

The second phase reference used during measurements with a mixer-based NVNA recovers the true phases of measured waves after the local oscillator frequency changes. This phase reference does not contribute systematic error because the absolute phase values are mathematically cancelled during calibration (when measuring the first phase reference) and therefore they do not need to be known. Random errors due to connector and cable repeatability are also insignificant when measuring this device because their contributions are also cancelled during calibration and the device is never reconnected or moved afterwards. Instead, for this phase reference the dominant uncertainty contribution is drift, because to quantify the varying local oscillator phase and correct the DUT measurements, the reference phase must be constant. For this reason commercially available phase references will typically use temperature-controlled electronics to minimise the drift of the produced harmonic tones, which improves the issue [37]. Because this phase reference is driven by another source (usually from within the NVNA, e.g. a 10 MHz reference clock), the phase noise and amplitude of this source will also have a significant effect on the uncertainty contribution. Blockley *et al.* provide an excellent review of this source of uncertainty and it's impact on NVNA measurements [38]. The key finding of this review is that for measurements using an IFBW of 100 Hz or below, the uncertainty contribution of random errors originating from the phase reference is negligible for typical DUT measurements. Although for VNA measurements IFBW settings above 100 Hz can be common for wideband frequency sweeps, the harmonic nature of NVNA measurements means that the IFBW is typically 30 Hz or below, which satisfies the recommendation from [38]

To provide a suitably accurate characterisation of the phase reference a sampling oscilloscope can be used to measure the output of the device [39]. This oscilloscope can be calibrated using electro-optic techniques traceable to primary standards [40], [41]. This service is provided by NIST for Keysight phase references and includes information on the uncertainty in the amplitude and phase of each measured harmonic [42]. The covariances between these quantities are not included, however, which can significantly affect the combined uncertainty of some nonlinear DUT measurements, including those involving cross-frequency terms such as behavioural models described later in Chapter 5.

Therefore, for this project two phase references were taken to NIST and re-characterised. The results of this characterisation included covariance information between the tone measurements at different harmonics and therefore can be used in a rigorous evaluation of uncertainty for nonlinear behavioural models. Results of the characterisation are shown in Figures 4.8 and 4.9. Due to the

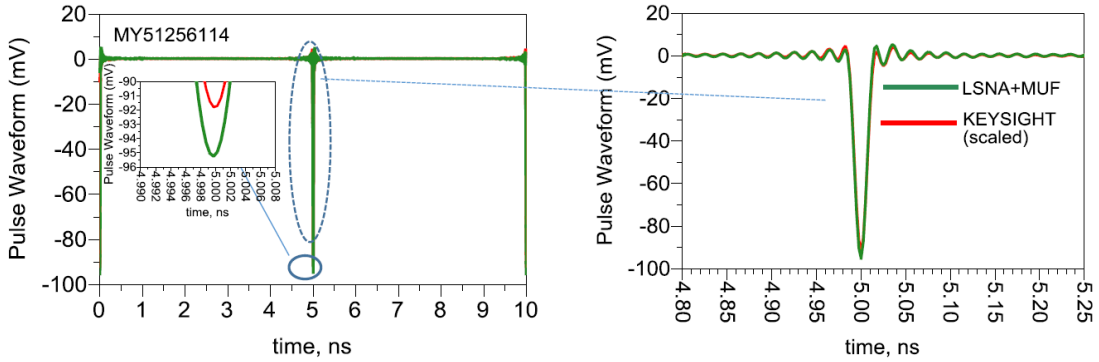


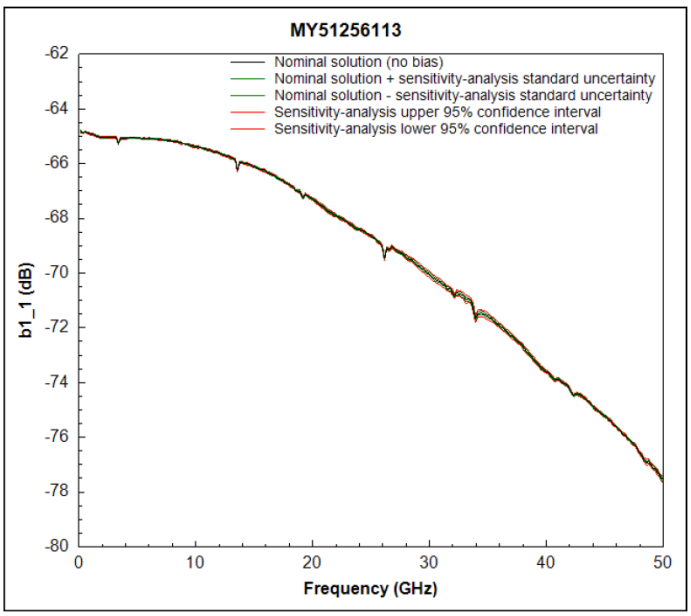
Figure 4.8: Time domain representation of the pulse train which creates the harmonic comb in the frequency domain. The pulses repeat every 5 ns when the generator is driven with a 10 MHz stimulus. This figure is used with permission from a report produced by Gustavo Avolio, who performed the measurements at NIST. The results show good agreement between the Keysight characterisation stored on the device memory.

considerable effort required to perform a characterisation using an electro-optically calibrated sampling oscilloscope, a new technique was trialled to characterise our phase references using an NVNA. The instrument was calibrated using a phase reference which already had a traceable characterisation including covariance information, from which the our phase references were then measured. This will produce a characterisation with larger uncertainties because an additional measurement has been included in the traceability chain. The results show uncertainties of around 0.5 degrees across the spectrum, with a considerable ripple in measured phase occurring above 35 GHz. The origin of this ripple was not found during the course of the project but occurred during de-embedding steps.

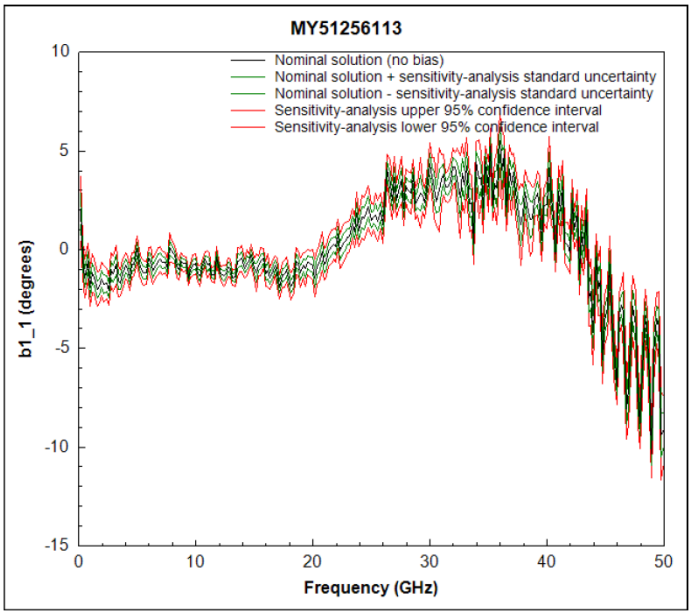
4.5.2 Power Meter

Power measurement is a well-established field of electromagnetic wave metrology and is widely applied to measurements in industry as well as academia. Because of this, manufacturers have produced application notes and software to help users evaluate the uncertainty in their power meter measurements [42], [43]. The uncertainty evaluation presented in this project is based on the power meter measurement model used by Keysight and described in [42], due to the use of Keysight power meters and the fact that this model is already included in the MUF software.

The distribution of uncertainty contributions are shown in Figure 4.10. The impedance



(a)



(b)

Figure 4.9: Results of the phase reference characterisation performed at NIST for unit MY51256113 displayed using the MUF measurement viewer software. The device was connected to port one of the NVNA performing the measurement, so the amplitude (a) and phase (b) of the b wave for the fundamental harmonic at port one (b1_1) is shown.

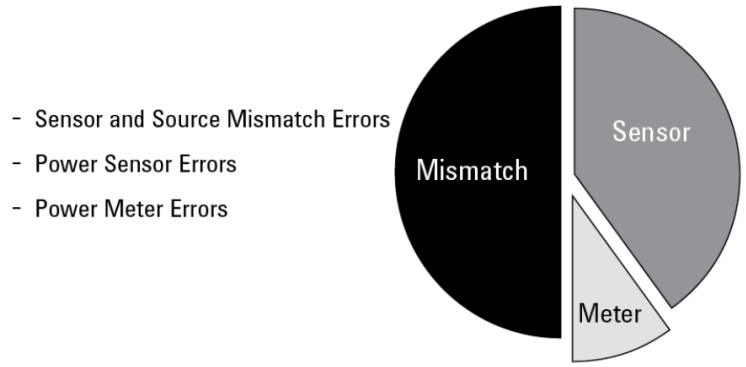


Figure 4.10: This chart shows a typical distribution of uncertainty values for its three largest causes: mismatch, sensor and meter specifications. It reveals why low SWR specifications for the power sensor and source is so crucial [42].

mismatch between the source to be measured and the power sensor is the dominant source of uncertainty, which is not surprising as it has a direct effect on the amount of power that reaches the sensing element.

The measurement model for the power meter can be written as

$$P = \frac{M_{\mathrm{u}}(P_{\mathrm{m}} - t)}{K_{\mathrm{b}} m}, \tag{4.9}$$

where P is the power incident to the sensor (the measurand), M_{u} is the gain due to mismatch of the power sensor, P_{m} is the value given by the power meter, t is the sum of offset errors including noise and zeroing errors, K_{b} is the calibration factor of the power sensor, and m is the sum of multiplicative errors including sensor calibration and instrumentation errors.

For an ideal measurement, M_{u} and m are all equal to one and t is equal to zero. When the model is used with a simple evaluation of uncertainty, the worst-case values for each error can be inserted or the root-sum-of-squares (RSS) calculated to provide a range of values which the result could take. This is the propagation technique suggested in the [42]. The MUF implementation instead sets all the input quantities to their nominal values (e.g. one or zero for the pure error terms) and instead assigns distributions to each of them. The uncertainty propagation then samples each input quantity from their distribution and obtains a sample for the measurand. A sensitivity analysis can also be performed for this model, and an example result is shown in Figure 4.11.

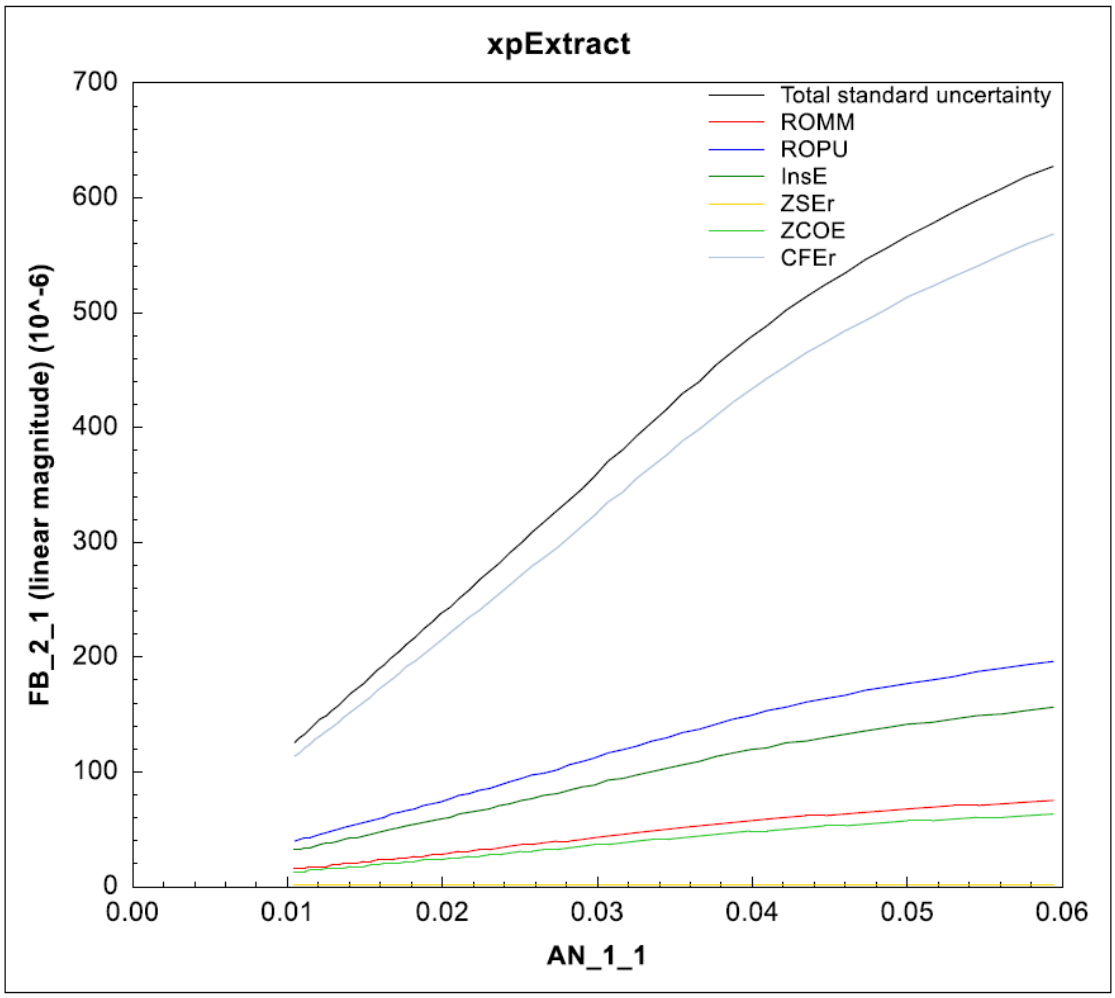


Figure 4.11: An example sensitivity analysis showing the effect of power meter uncertainties on an X-parameter measurement. The plot shows the amplitude of the X-parameter representing the large-signal output of the amplifier into a matched load versus a swept input power. The values are in units of square-root Watts. The input quantities ROMM, ROPU and InsE all contribute to m, ZSEr and ZCOE contribute to t, and CFEr represents the error in K_{b}. Mismatch error is not included in the power meter model but is modelled as a cascaded adapter attached to the power sensor instead.

4.5.3 Propagation of Uncertainties

Although there are many guides and software frameworks available for propagating uncertainty through VNA measurements, there are only two published options currently available for NVNA measurements [31], [44].

4.5.3.1 Analytical Covariance-Based Propagation

Lin and Zhang published a covariance-based analytical method for linear propagation of uncertainty which allows a GUM-based technique to be applied to NVNA uncertainties [44]. Partial derivates with respect to all input quantities have been derived for the NVNA measurement model, which is shown graphically in Figure 4.12. These include the use of Cauchy-Riemann derivatives in order to accommodate ratios of the complex components used to calculate magnitude and phase values for the respective parts of the absolute calibrations.

4.5.3.2 Numerical Propagation

The MUF includes both s-parameter and wave-parameter measurement models for a VNA. The wave-parameter model was extended at NIST and used to create an NVNA measurement model which can be used to propagate uncertainty for nonlinear device measurements using the numerical methods included in the MUF - namely Monte Carlo and sequential perturbation. This was the approach chosen for the work in this project because there were already published results from this framework and it was actively supported [45]. In addition, the MUF includes preserves information about correlations between frequencies in the form of indexed Monte Carlo samples. This is an alternative to the very large covariance matrices required to otherwise store this information at each step of the propagation (although there have been recent efforts to improve this burden using principal component compression [46]). This ability is significant when using the results of the measurement to produce time-domain or cross-frequency information as required by nonlinear behavioural models, where without these correlations included the combined uncertainty may be considerably perturbed.

Figure shows a screenshot of the MUF software **VNA Uncertainty Calculator** with the LSNA tab open. This part of the software allows a user to enter models for the power meter and adapters, and measurements for the power and phase calibrations. The phase reference does not use a model, but instead a characterisation file including covariances as described earlier in this section.

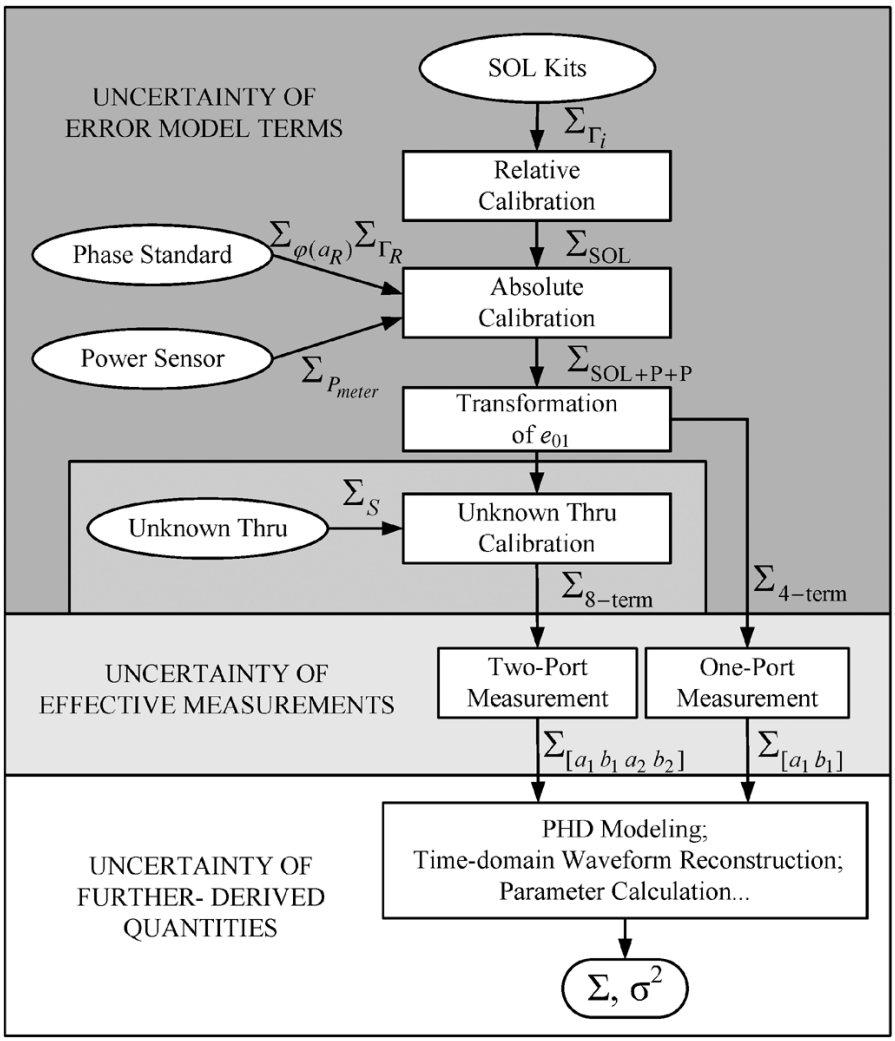


Figure 4.12: The NVNA measurement model used in [44].

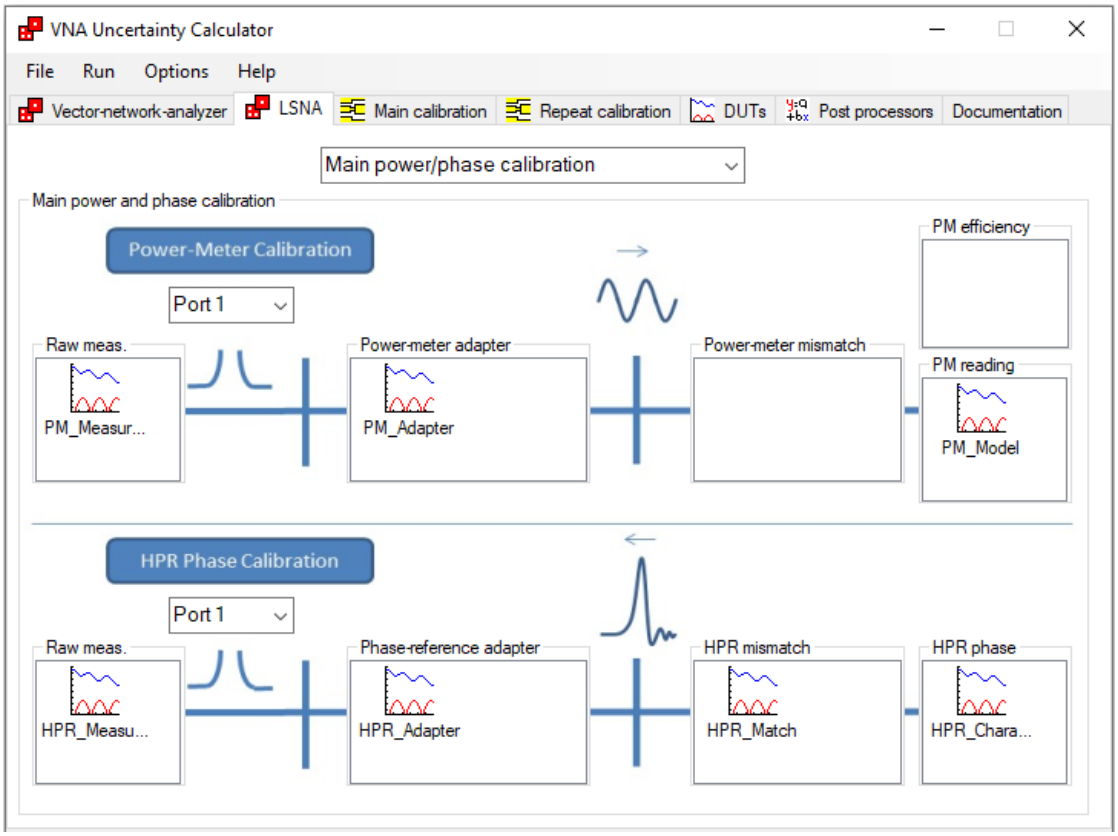


Figure 4.13: The LSNA calibration tab of the MUF **VNA Uncertainty Calculator** application. Each file icon represents a ".meas" file, which contains an index of Monte Carlo samples generated from a statistical analysis of repeated measurements for each quantity. It is by this method that the MUF propagates distribution information and preserves covariance information. The power meter mismatch and efficiency input boxes are empty as they are included in the model which generated the processed power meter reading file ("PM_Model").

4.6 Conclusion

This chapter has studied the application of uncertainty evaluation to measurements performed on both VNA and NVNA instruments. After a brief introduction to typical sources of uncertainty encountered in these measurements, two versions of VNA measurement models were explained. The residual error model provides a less computationally-intensive propagation of uncertainty, which was noticeably beneficial in the years when it was developed, but this concern is now less relevant. It was shown that the residual error model can also suffer from problems during the ripple measurement method used to characterise some of the input quantities, and is also not reliable when used in submillimetre-wave rectangular metallic waveguide. In contrast, the rigorous VNA measurement model provides more accurate values for combined uncertainties, especially as it correctly preserves correlations through the propagation. This can be achieved either via linear propagation methods as defined in the GUM document, or via Monte Carlo techniques as defined in the GUM supplement, both of which were explained in Chapter 3.

The extension of VNA uncertainty to NVNA measurements was also discussed, where there is a requirement to add the absolute power and phase calibrations to the VNA measurement model. Due to NVNA uncertainty evaluation being relatively young in the field of microwave metrology, there are only two available frameworks to provide such evaluations. The single software solution, the MUF, provides multiple propagation methods and includes a sensitivity analysis feature. It can propagate cross-frequency correlations which are important for use in nonlinear behavioural models. For these reasons, the MUF was selected for use in this project as the framework to use to propagate measurement uncertainty into nonlinear behavioural models.

References

[1] International Organization for Standardization, *ISO 9000: International standards for quality management*, Genève, Switzerland: International Organization for Standardization, 2015. [Online]. Available: `https://www.iso.org`.

[2] ——, *ISO 17025: General requirements for the competence of testing and calibration laboratories*, Genève, Switzerland: International Organization for Standardization, 2005. [Online]. Available: `https://www.iso.org`.

[3] International Bureau of Weights and Measures (BIPM), *The International System of Units (SI)*, 8th ed. 2019. [Online]. Available: `https://www.bipm.org`.

[4] ——, *A concise summary of the International System of Units, the SI*, 8th ed. 2019. [Online]. Available: `https://www.bipm.org`.

[5] R. N. Clarke, J. P. Ide, G. R. Orford, and N. M. Ridler, "Draft EAL procedure for the assessment of vector network analysers (VNA), ANAMET report 002," National Physical Laboratory, Tech. Rep. 002, Sep. 1996, Available from the National Physical Laboratory.

[6] J. P. M. de Vreede, "Draft procedure for the assessment of vector network analysers (VNA), ANAMET report 019," National Physical Laboratory, Tech. Rep. 019, Oct. 1998, Available from the National Physical Laboratory.

[7] ——, "Draft EA guidance document: Assessment of calibrated vector network analysers (VNA), ANAMET report 026," National Physical Laboratory, Tech. Rep. 026, Aug. 1999, Available from the National Physical Laboratory.

[8] European co-operation for Accreditation, "EA guidelines on the evaluation of vector network analysers (VNA)," Tech. Rep. EA-10/12, May 2000.

[9] European Association of National Metrology Institutes (EURAMET), "Guidelines on the evaluation of vector network analysers (VNA), version 2," Tech. Rep. cg-12, Mar. 2011.

[10] M. Hiebel, *Fundamentals of Vector Network Analysis*. Rohde & Schwarz, 2008, ISBN: 9783939837060.

[11] Keysight Technologies. Downloadable vector network analyzer uncertainty calculator, [Online]. Available: `https://www.keysight.com/main/software.jspx?ckey=1000000418:epsg:sud&cc=GB` (visited on 07/16/2019).

[12] Anritsu. Exact uncertainty calculator, [Online]. Available: `https://www.anritsu.com/en-GB/test-measurement/support/downloads/software/dwl19572` (visited on 07/16/2019).

[13] Keysight Technologies, "Specifying calibration standards and kits for keysight vector network analyzers," Tech. Rep. 1GP43_0E, Jul. 2016.

[14] A. Lewandowski, "Multi-frequency approach to vector-network-analyzer scattering-parameter measurements," PhD thesis, Warsaw University of Technology, 2010.

[15] D. K. Rytting, "Network analyzer accuracy overview," in *58th ARFTG Conference Digest*, vol. 40, Nov. 2001, pp. 1–13. DOI: `10.1109/ARFTG.2001.327486`.

[16] J. Martens, "On quantifying the effects of receiver linearity on VNA calibrations," in *2007 70th ARFTG Microwave Measurement Conference (ARFTG)*, Nov. 2007, pp. 1–6. DOI: `10.1109/ARFTG.2007.8376223`.

[17] L. Stant, P. Aaen, and N. Ridler, "Evaluating residual errors in waveguide network analysers from microwave to submillimetre-wave frequencies," in *IET Colloquium on Millimetre-Wave and Terahertz Engineering & Technology 2016*, Institution of Engineering and Technology (IET), 2016. DOI: `10.1049/ic.2016.0016`.

[18] ——, "Evaluating residual errors in waveguide VNAs from microwave to submillimetre-wave frequencies," *IET Microwaves Antennas Propag.*, vol. 11, no. 3, pp. 324–329, Feb. 2017. DOI: `10.1049/iet-map.2016.0455`.

[19] D. Bannister, E. Griffin, and T. Hodgetts, *On the dimensional tolerances of rectangular waveguide for reflectometry at millimetric wavelengths*. National Physical Laboratory, 1989.

[20] A. Kerr, E. Wollack, and N. Horner, "Waveguide flanges for ALMA instrumentation," *ALMA Memo*, vol. 278, 1999.

[21] A. Kerr, "Mismatch caused by waveguide tolerances, corner radii, and flange misalignment," *Electronics Div. Tech. Note*, vol. 215, 2010.

[22] D. F. Williams, "500 ghz–750 ghz rectangular-waveguide vector-network-analyzer calibrations," *IEEE Transactions on Terahertz Science and Technology*, vol. 1, no. 2, pp. 364–377, 2011.

[23] H. Li, A. R. Kerr, J. L. Hesler, and R. M. Weikle, "Repeatability of waveguide flanges with worst-case tolerances in the 500–750 ghz band," in *79th ARFTG Microwave Measurement Conference*, IEEE, 2012, pp. 1–8.

[24] "IEEE standard for rectangular metallic waveguides and their interfaces for frequencies of 110 ghz and above–part 2: Waveguide interfaces," *IEEE Std 1785.2-2016*, pp. 1–22, Sep. 2016. DOI: 10.1109/IEEESTD.2016.7564020.

[25] A. Arsenovic, R. M. Weikle, and J. Hesler, "Two port calibration insensitive to flange misalignment," in *84th ARFTG Microwave Measurement Conference*, Dec. 2014, pp. 1–7. DOI: 10.1109/ARFTG.2014.7013416.

[26] M. Garelli and A. Ferrero, "A unified theory for *s*-parameter uncertainty evaluation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 3844–3855, Dec. 2012, ISSN: 0018-9480. DOI: 10.1109/TMTT.2012.2221733.

[27] Keysight Technologies, "VNA measurement uncertainty," Tech. Rep. 5992-3595EN, Jun. 2019.

[28] METAS. VNA tools II, [Online]. Available: https://www.metas.ch/metas/en/home/fabe/hochfrequenz/vna-tools.html (visited on 07/16/2019).

[29] ——, Metas unclib, [Online]. Available: https://www.metas.ch/metas/en/home/fabe/hochfrequenz/unclib.html (visited on 07/16/2019).

[30] M. Wollensack, J. Hoffmann, J. Ruefenacht, and M. Zeier, "VNA tools II: S-parameter uncertainty calculation," in *79th ARFTG Microwave Measurement Conference*, Jun. 2012, pp. 1–5. DOI: 10.1109/ARFTG79.2012.6291183.

[31] Keysight Technologies. On-wafer calibration software, [Online]. Available: https://www.nist.gov/services-resources/software/wafer-calibration-software (visited on 07/16/2019).

[32] R. A. Ginley, "Kicking the tires of the nist microwave uncertainty framework, part 1," in *2016 88th ARFTG Microwave Measurement Conference (ARFTG)*, Dec. 2016, pp. 1–4. DOI: 10.1109/ARFTG.2016.7839722.

[33] ——, "Kicking the tires of the nist microwave uncertainty framework, part 2," in *2017 90th ARFTG Microwave Measurement Symposium (ARFTG)*, Nov. 2017, pp. 1–4. DOI: 10.1109/ARFTG.2017.8255877.

[34] J. A. Jargon, D. F. Williams, and A. Sanders, "Three-port vector-network-analyzer calibrations using the nist microwave uncertainty framework," in *2019 92nd ARFTG Microwave Measurement Conference (ARFTG)*, Jan. 2019, pp. 1–4. DOI: 10.1109/ARFTG.2019.8637226.

[35] J. N. H. Dortmans, J. T. Quimby, K. A. Remley, D. F. Williams, J. Senic, R. Sun, and P. B. Papazian, "Design of a portable verification artifact for millimeter-wave-frequency channel sounders1," *IEEE Transactions on Antennas and Propagation*, pp. 1–1, 2019, ISSN: 0018-926X. DOI: 10.1109/TAP.2019.2902623.

[36] D. Gu, X. Lu, B. F. Jamroz, D. F. Williams, X. Cui, and A. W. Sanders, "NIST-traceable microwave power measurement in a waveguide calorimeter with correlated uncertainties," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 6, pp. 2280–2287, Jun. 2019, ISSN: 0018-9456. DOI: 10.1109/TIM.2018.2886731.

[37] J. A. Jargon, D. C. DeGroot, and D. F. Vecchia, "Repeatability study of commercial harmonic phase standards measured by a nonlinear vector network analyzer," in *Conference, 2003. Fall 2003. 62nd ARFTG Microwave Measurements*, Dec. 2003, pp. 243–258. DOI: 10.1109/ARFTGF.2003.1459783.

[38] P. S. Blockley, J. B. Scott, D. Gunyan, and A. E. Parker, "The random component of mixer-based nonlinear vector network analyzer measurement uncertainty," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 10, pp. 2231–2239, Oct. 2007, ISSN: 0018-9480. DOI: 10.1109/TMTT.2007.906515.

[39] H. C. Reader, D. F. Williams, P. D. Hale, and T. S. Clement, "Comb-generator characterization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 2, pp. 515–521, Feb. 2008, ISSN: 0018-9480. DOI: 10.1109/TMTT.2007.914630.

[40] D. F. Williams, A. Lewandowski, T. S. Clement, J. C. M. Wang, P. D. Hale, J. M. Morgan, D. A. Keenan, and A. Dienstfrey, "Covariance-based uncertainty analysis of the nist electrooptic sampling system," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 1, pp. 481–491, Jan. 2006, ISSN: 0018-9480. DOI: 10.1109/TMTT.2005.860492.

[41] P. D. Hale, A. Dienstfrey, J. C. M. Wang, D. F. Williams, A. Lewandowski, D. A. Keenan, and T. S. Clement, "Traceable waveform calibration with a covariance-based uncertainty analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 10, pp. 3554–3568, Oct. 2009, ISSN: 0018-9456. DOI: 10.1109/TIM.2009.2018012.

[42] Keysight Technologies, "Fundamentals of RF and Microwave Power Measurements (Part 3)," Tech. Rep. 5989-7619EN, Sep. 2017.

[43] Rohde & Schwarz, "Program for measurement uncertainty analysis with Rohde & Schwarz power meters," Tech. Rep. 5989-4840EN, Jun. 2001.

[44] M. Lin and Y. Zhang, "Covariance-matrix-based uncertainty analysis for NVNA measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 1, pp. 93–102, Jan. 2012. DOI: 10.1109/tim.2011.2150650.

[45] G. Avolio, A. Raffo, J. Jargon, P. D. Hale, D. M. M. .-.-P. Schreurs, and D. F. Williams, "Evaluation of uncertainty in temporal waveforms of microwave transistors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 7, pp. 2353–2363, Jul. 2015, ISSN: 0018-9480. DOI: 10.1109/TMTT.2015.2432765.

[46] D. A. Humphreys, P. M. Harris, M. Rodríguez-Higuero, F. A. Mubarak, D. Zhao, and K. Ojasalo, "Principal component compression method for covariance matrices used for uncertainty propagation," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 2, pp. 356–365, Feb. 2015, ISSN: 0018-9456. DOI: 10.1109/TIM.2014.2340640.

5 Propagating Measurement Uncertainty into Nonlinear Behavioural Models

5.1 Introduction

5.2 The X-Parameter Model

5.2.1 Model Definition

5.2.2 Extraction Procedure

5.2.3 Applications of X-Parameters

5.3 Design and Simulation using Nonlinear Behavioural Models Incorporating Measurement Uncertainty

5.4 Conclusions

Testing, testing[1], [2].

Bibliography

[1] L. Stant, P. Aaen, and N. Ridler, "Evaluating residual errors in waveguide network analysers from microwave to submillimetre-wave frequencies," in *IET Colloquium on Millimetre-Wave and Terahertz Engineering & Technology 2016*, Institution of Engineering and Technology (IET), 2016. DOI: 10.1049/ic.2016.0016.

[2] ——, "Comparing methods for evaluating measurement uncertainty given in the JCGM 'evaluation of measurement data' documents," *Measurement*, vol. 94, pp. 847–851, Dec. 2016. DOI: 10.1016/j.measurement.2016.08.015.

6 Conclusions

6.1 Further Work

Testing, testing[1], [2].

Bibliography

[1] L. Stant, P. Aaen, and N. Ridler, "Evaluating residual errors in waveguide network analysers from microwave to submillimetre-wave frequencies," in *IET Colloquium on Millimetre-Wave and Terahertz Engineering & Technology 2016*, Institution of Engineering and Technology (IET), 2016. DOI: 10.1049/ic.2016.0016.

[2] ——, "Comparing methods for evaluating measurement uncertainty given in the JCGM 'evaluation of measurement data' documents," *Measurement*, vol. 94, pp. 847–851, Dec. 2016. DOI: 10.1016/j.measurement.2016.08.015.