Developments on the Forefront of Terahertz Metrology and Calibration

A Dissertation

Presented to

the Faculty of the School of Engineering and Applied Science

UNIVERSITY of VIRGINIA

In Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy of Electrical Engineering

by

Alex Arsenovic

10/10/2012

APPROVAL SHEET

The thesis

is submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

AUTHOR

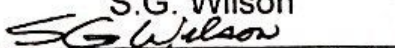The thesis has been read and approved by the examining committee:

Robert Weikle

Advisor

N. Scott Barker

A. R. Kerr

A. W. Lichtenberger

S.G. Wilson

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

December
2012

Abstract

The terahertz spectrum, often defined as the 300GHz-3THz frequency band, is a relatively unexplored segment of the electromagnetic spectrum, lying between the microwave and far infrared regions. Although measurement capabilities at these frequencies are just now being realized, there are already numerous applications such as imaging, spectroscopy, radio astronomy, and medical diagnostics[1]. To support this growing industry, a proper measurement infrastructure is needed, similar to what is currently available at 100 GHz and below. Efforts to develop the hardware[2, 3] and components[4, 5] needed to realize this measurement infrastructure are being made. Complementary to these advances in new hardware, specialized measurement procedures and improved calibration algorithms are being created[6, 7]. This work presents contributions to the field of terahertz metrology, specifically to the area of calibration and uncertainty analysis.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Motivation

1.1.1. Why THz Metrology?

> Numerous applications are beginning to emerge in the terahertz fre-
> quency region, creating a demand for a measurement infrastructure sim-
> ilar to what is currently available at 100GHz and below.

The terahertz spectrum, often defined as 300GHz-3THz frequency band, is a rel-
atively unexplored segment of the electromagnetic spectrum, lying between the
microwave and far infrared regions. Although measurement capabilities at these
frequencies are just now being realized, there are already numerous applications
such as imaging, spectroscopy, radio astronomy, and medical diagnostics[1]. The
rising level of interest is reflected in the recent creation of a new journal focusing
on the subject, the *IEEE Transactions on Terahertz Science and Technology*. To
support this growing industry, a proper measurement infrastructure is needed. This
infrastructure consists of measurement instruments, components, and specialized
procedures.

The essential piece of hardware needed to characterize devices at microwave frequencies is the Vector Network Analyzer (VNA). Recently, vector measurements have been extended to 1.1 THz by Virginia Diodes[8]. A picture of the VDI frequency extender system in conjunction with an Aglient VNA, is shown in Fig. 1.1. This new capability has created a demand for other measurement hardware that can operate at terahertz frequencies. One such component is the on-wafer probe, which allows micro-fabricated devices to be characterized without manual assembly into a waveguide test block. Terahertz on-wafer probes operating at frequencies up to 750GHz have recently been developed here at the University of Virginia [4, 5], broadening measurement capabilities to an entire new medium of devices.
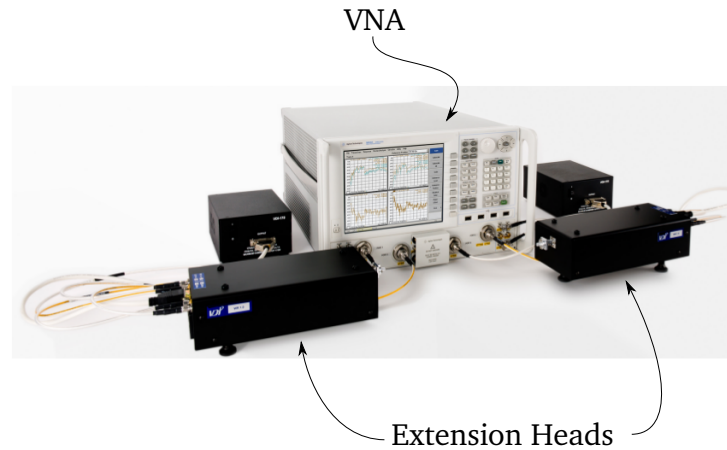
Figure 1.1.: VNA and frequency extension heads from VDI.

Another area of focus is the design and specification of waveguide flanges used to connect rectangular waveguide components [9]. As the operating frequencies increase, waveguide misalignment caused by mechanical tolerances becomes a severe problem. Not only does misalignment produce reflections at each flange connection, but it creates measurement inaccuracies by inducing calibration error[10]. This

problem is being addressed by the IEEE P1875 working group[2, 3, 11]. Other improvements to the terahertz measurement infrastructure are being made in the form of specialized measurement procedures and improved calibration algorithms[6, 7]. Along with these new algorithms, new calibration standards are being explored[12], and design variations of traditional standards are being evaluated[10].

The demand for higher frequency systems continues to produce new challenges for metrologists. In order to solve many of these measurement challenges, the fundamentals of calibration must be understood.

1.1.2. Why Calibration?

> Due to the complexity and subtitles of microwave systems, achieving accurate measurements in novel environments requires in depth knowledge of calibration theory.

Scattering parameter measurements are typically made with a Vector Network Analyzer (VNA). As its name suggests, this instrument provides *vector* measurements, meaning both amplitude and phase information of the transmitted and reflected electromagnetic waves is available. Before useful measurements can be made with a VNA, the system must be *calibrated* by measuring a set of characterized *standards*. The purpose of calibration is to remove the unwanted systematic electrical effects of intermediary circuitry existing between the internals of the VNA and the test ports. Because the system is vectorial in nature, the theory of calibration is significantly more involved than with scalar power-detection systems.

A calibration will always be imperfect due to the numerous sources of error and uncertainties. These imperfections will create inaccuracy in corrected measurements, that may or may not be perceivable. At lower frequencies, or in conventional mea-

surement environments, well known calibration procedures and accurately characterized calibration standards exist, so that achieving and verifying a high quality calibration is relatively straightforward. At terahertz frequencies, however, well-tested calibration procedures and standards are generally unavailable. This makes creating a good calibration much more challenging. Consequently, knowledge of calibration theory becomes essential for the experimentalist working at these frequencies.

Although the main application of this work is terahertz metrology, similar challenges exist in non-conventional measurements at lower frequencies. Research environments frequently require custom experimental setups to make specific, and sometimes single-use measurements. Examples of two scenarios are given in sec. 4.4.3 and sec. A.5. In such cases, new measurement procedures are created instead of followed, making calibration as difficult as it is in the terahertz region.

1.1.3. The Audience

This thesis is written for an audience with a background in microwave engineering, but not necessarily metrology. To make the content accessible, chapter 2 provides the necessary fundamentals of calibration theory. The most direct audience for this work is the academic community, RF industry, and research laboratories that are working to build the terahertz measurement infrastructure, or use terahertz measurements for material and device characterization. A partial list of who the author identifies as the immediate audience for this work is given in Tab. 1.1 below. This work is also of interest to the more general audience of microwave experimentalists who deal with non-conventional, research type measurement setups.

National Laboratories	Companies
NASA Jet Propulsion Laboratory (JPL)	Virginia Diodes
National Institute of Standards and Technology (NIST)	Northrup Grumman
National Physical Laboratory (NPL)	Agilent
Fraunhofer Institute	Cascade microtech
National Radio Astronomy Observatory (NRAO)	GGB Industries
	Dominion Mircoprobes

Table 1.1.: Companies and research laboratories that are involved with terahertz metrology.

1.2. Outline of Content

Each chapter of this thesis is intended to be a self-contained entity. To accomplish this, a small amount of theory is repeated at the beginning of a few chapters. I believe the modularity is worth the redundancy. The content begins with chapter 2, which provides the necessary foundations of calibration theory so that the subsequent content may be appreciated. The delivery is somewhat unique, with emphasis on basic concepts as opposed to specific calibration algorithms. A detailed description of the calibration processing chain is given, identifying specific errors effecting corrected measurements. Discussion of these errors introduces the topic of uncertainty analysis and error metrics. The chapter closes with an optional section on the Mobius Transformation, which provides a slightly more abstract mathematical perspective.

The next two chapters contain specific contributions to the field of terahertz metrology. In chapter 3, a general calibration technique is presented that utilizes partially known calibrations standards. It is emphasized that although this technique was motivated by the lack of well characterized standards at terahertz frequencies, it can be used in other non-conventional measurement environments as well.

The following chapter, chapter 4, presents an experimentally based method of cali-

bration uncertainty analysis. Demonstrated by numerous examples, the procedure is conceptually simple and produces intuitive results. Although formulated for s-parameter measurements, the basic method can be extended to any measurement system.

The appendices contain a mixture of different work. Some of the sections contain practical applications of the two calibration concepts just described, such as sec. A.6 and sec. A.5. Others are tangentially related applications, such as those in sec. E.1 and sec. A.1. The content in the Appendix is less refined than that in the main body, but has been included for documentation purposes, and as a foundation for potential new areas of work.

1.3. Reflection on Work

Most, if not all of the techniques developed in this thesis were born out of the specific needs of our research group. This accounts for the sporadic nature of the appendices. Once a technique was deployed in the laboratory and proven to be useful, it was revisited more formally and generalized to suit a broader scope of problems. In general, the University of Virginia millimeter wave research lab has proved to be a very prolific generator of interesting problems, and I am thankful for the opportunity to be a part of it.

A note on nomenclature.

With the design of mathematical notation there is a compromise between brevity of presentation and unambiguity of meaning. The problem is further complicated by conventions used in the literature, which are frequently unclear and inconsistent. Although the notation in this document doesn't follow any strict convention, I have

tried to emphasize quantities of importance with bold highlighting, and given different objects different kinds of symbols; vectors are lowercase, matrices are upper case, etc.

One source of confusion is matrix indices. A common convention when dealing with a series of networks is to let the indices denote the port numbers. So a 2x2 s-parameter matrix could have entries $s_{11}, s_{12}, \dots$, or $s_{00}, s_{01}, \dots$, depending on the number given to either port. Frequently, the matrices are pulled out of the original context, and the indexing scheme can appear arbitrary and confusing. In either case, it is not a substantial issue as long as one is aware of it.

2. Calibration Theory

2.1. Introduction

2.1.1. Outline

This chapter provides the theoretical foundations of calibration theory necessary to appreciate the work presented in this thesis. It should be accessible to those with basic knowledge of microwave network theory. Following a very brief history, the theory begins in sec. 2.2. This section contains a basic explanation of unterminating and de-embedding, scattering transfer parameters, tiered calibration, and details of the calibration processing chain. The following section, sec. 2.3, discusses calibration uncertainty, residual error analysis, and biased and unbiased error metrics. The final section, sec. 2.4, gives a more abstract mathematical perspective of calibration, focusing on properties of the Mobius transformation. Although the author believes such a perspective is of great value, it is somewhat unconventional, and therefore the final section is included for those interested in the topic.

2.1.2. A (very) Brief History of Calibration

Before delving into calibration theory, a short history of microwave measurements is in order. Much of the history presented here is summarized in Section II of [13].

One of the first instruments used to make reflection coefficient measurements was the slotted line. This instrument, shown in Fig. 2.1, allows a probe to slide along a section of uniform transmission line or waveguide, sampling the power at the different electrical delays from the test port. Assuming that the effect of the probe is negligible, the measured power is a linear combination of incident and reflected waves. Given a sufficient number of measurements at the appropriate delays, the reflection coefficient magnitude and phase can be determined.
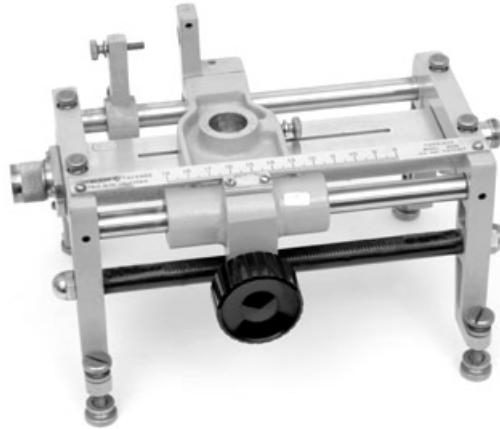
Figure 2.1.: Picture of a slotted line (from Google)

The drawback to the slotted line, and other similar instruments, is the measurement accuracy depends on manufacturing precision, such as the geometry of probe, sliding mechanism, etc. As a partial solution to this problem, the tuned reflectometer was introduced. As it's name suggests, the tuned reflectometer could be *tuned* by the metrologist to compensate for manufacturing imperfections. This was, however, a very time-consuming and laborious task, which required a skilled operator to produce accurate measurements[13].

With the advent of the Automatic Network Analyzer (ANA), which we refer to to-day as the Vector Network Analyzer (VNA), a major paradigm shift in measurement strategy occurred. Instead of attempting to remove imperfections by using more sophisticated manufacturing techniques, the imperfections were explicitly acknowledged and removed through mathematical means. This technique is now known as *calibration.*

In other words,

> *attempts to produce perfect hardware were replaced by the intelligent use of imperfect hardware* [13].

2.2. Calibration

2.2.1. Unterminating and De-embedding

Calibration is best understood as a specific case of the more general procedure known as *unterminating and de-embedding.* This chapter gives a description of the procedure, following closely the perspective of Bauer and Penfield[14].

At microwave frequencies it is generally not possible to measure quantities of interest, such as reflection coefficient, at the location of interest. Instead, it is usually possible to measure the quantity of interest, or a related quantity, at some other location. For example, one may wish to measure the reflection coefficient of a device, but that device is electrically far away from the measurement instrument. This scenario is illustrated below in Fig. 2.2. In the figure, the actual reflection coefficient (**a**) of the device under test (DUT) is said to be *embedded* behind the intervening two port network, producing a measurable reflection coefficient (**m**). The goal is to remove the effects of the embedding network from the measurement of (**m**) so that the DUT

can be characterized.

The procedure to accomplish this is broken up into two parts, referred to as *unter-minating* and *de-embedding*. The determination of the embedding network is known as *unterminating*. The subsequent removal of the electrical effects of the embedding network on a device is known as *de-embedding*. In this context, calibration can be described as unterminating the effects of a measurement system by measuring a set of known devices, and then de-embedding those effects from future measurements.
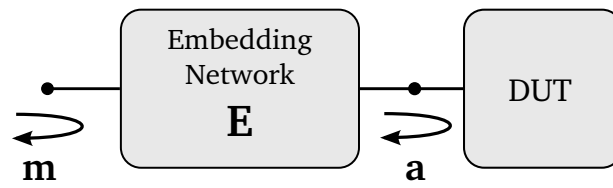
Figure 2.2.: Diagram of an embedded device.

Assuming all the intervening circuitry between the measurement reference plane and the device is linear, then it can be represented as a single two-port network, a result of linear network theory[15]. Another well known result from network theory is the relationship between the actual (true) reflection coefficient of a device ($\mathbf{a}$), the measured reflection coefficient ($\mathbf{m}$), and the embedding network ($\mathbf{E}$), given by the Möbius transformation.

$$\mathbf{m} = f\left(\mathbf{a}\right) = e_{00} + \frac{e_{01}e_{10} \cdot \mathbf{a}}{1 - e_{11} \cdot \mathbf{a}} \tag{2.1}$$

where

$$\mathbf{m} = \text{measured response}$$

$$\mathbf{a} = \text{actual response}$$

$$E = \begin{bmatrix} e_{00} & e_{01} \\ e_{10} & e_{11} \end{bmatrix} = \text{embedding network S-parameters}$$

This algebraic relationship can be expressed pictorially using a signal flow graph[16], as shown in Fig. 2.3.
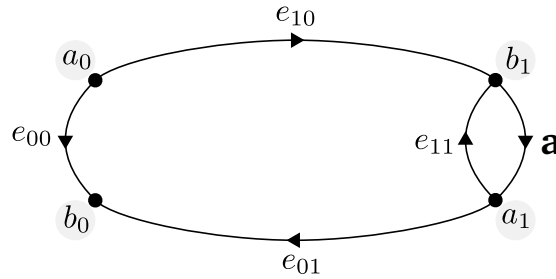
Figure 2.3.: Signal Flow Graph for a Terminated Two-Port Network

This relationship defines a unique, non-linear map between two planes; the *a-plane* and the *m-plane*, meaning there is one-to-one correspondence between measured reflection coefficient and the actual reflection coefficient of the device. Building upon Fig. 2.2, the reference planes are illustrated at their location within the circuit in Fig. 2.4. In the context of calibration, the *a-plane* is usually the test-port of the measurement system and the *m-plane* is the raw measurement produced by the VNA. The goal is to determine the parameters of the transform, so that it may be inverted, thereby producing corrected measurements.
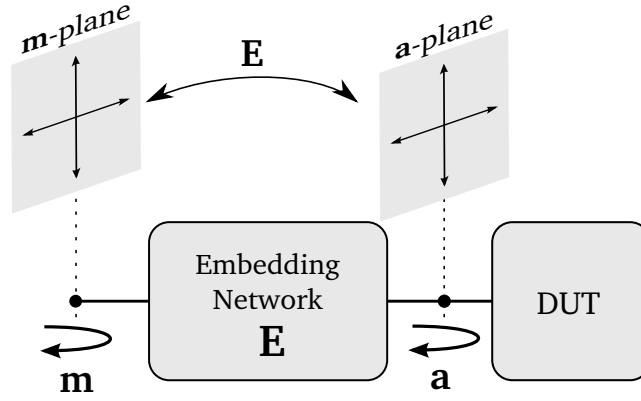
Figure 2.4.: Diagram of an embedded device illustrating the two reference planes of interest.

It is of practical importance to recognize that the term $(e_{10}e_{01})$ cannot be seperated, therefore the terms (e_{10}) and (e_{01}) cannot be found individually. This can be seen from equation (2.1), and makes physical sense. However, if the determined parameters are used solely for de-embedding, this is not of concern because the two terms never appear in isolation. Otherwise, if the values of (E) are of interest, then (E) can be completely solved for *only* if it is a reciprocal network $(e_{10} = e_{01})$ and the transmission phase is known within a factor of π [1].

Unterminating The electrical properties of the embedding network are determined by measuring a set of known devices at the test port. Such devices are called calibration standards. Mathematically, this means that the parameters of the transform are determined by observing how a set of known points in the a-plane is transformed on the m-plane. These parameters can be uniquely determined given three measurements. Because (2.1) defines a non-linear relationship between (**m**) and (**a**), an iterative algorithm must be used to determine the parameters of interest. Alternatively, (2.1) can be converted into a linear equation by treating the determinant of the embedding network's s-parameter matrix as a single variable ($\Delta e =$

---

[1]a sign ambiguity still exists because $\sqrt{e_{10}^2} = \pm e_{10}$

$e_{00}e_{11} - e_{01}e_{10})$ and mixing the independent and dependent variables to produce the following,

$$e_{11}\mathbf{m a} - \Delta e\,\mathbf{a} + e_{00} = \mathbf{m} \tag{2.2}$$

Given three raw measurements $(\mathbf{m}_{1\ldots 3})$ of calibration standards with known responses $(\mathbf{a}_{1\ldots 3})$, the unknowns $(e_{00}, e_{11}, \Delta e)$ can be directly found using matrix inversion.

$$\underbrace{\begin{bmatrix} \mathbf{m}_1 \cdot \mathbf{a}_1 & -\mathbf{a}_1 & 1 \\ \mathbf{m}_2 \cdot \mathbf{a}_2 & -\mathbf{a}_2 & 1 \\ \mathbf{m}_3 \cdot \mathbf{a}_3 & -\mathbf{a}_3 & 1 \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} e_{11} \\ \Delta e \\ e_{00} \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{m_1} \\ \mathbf{m_2} \\ \mathbf{m_3} \end{bmatrix}}_{\mathbf{b}} \tag{2.3}$$

$$\begin{aligned} \mathbf{A} \cdot \mathbf{x} &= \mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1} \cdot \mathbf{b} \end{aligned} \tag{2.4}$$

Given additional measurements, (2.3) becomes overdetermined. The least squares solution to such a system is found by calculating the pseudo-inverse of $\mathbf{A}$, given by $\mathbf{A}^{+} = (\mathbf{A}^{*}\mathbf{A})^{-1}\mathbf{A}^{*}$ [17]. Where the asterisk ($*$) represents conjugate transpose. Thus, (2.4) becomes,

$$\mathbf{x} \;=\; (\mathbf{A}^{*}\mathbf{A})^{-\mathbf{1}}\mathbf{A}^{*} \cdot \mathbf{b} \tag{2.5}$$

$$\;=\; \mathbf{A}^{+} \cdot \mathbf{b} \tag{2.6}$$

Over-determining the system allows for residual analysis, which is discussed in sec. 2.3. Other improvements, such as optimally weighted least squares have been discussed in the literature [18, 19].

De-embedding   Once the embedding network is characterized its effects can be removed from future measurements. This is done by inverting (2.1) ,

$$\mathbf{a} = \frac{\mathbf{m} - e_{00}}{e_{11} \cdot \mathbf{m} - \Delta e}$$

A convenient way to both conceptualize and implement unterminating and de-embedding is through scattering transfer parameters.

Scattering Transfer Parameters   It is well known that cascading two-port networks can be implemented with matrix multiplication by using what are known as *scattering transfer parameters,* or T-parameters[20]. These parameters are defined by linearly relating incoming and outgoing waves at each port. Given that the S-parameters are defined by

$$
\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} s_{00} & s_{01} \\ s_{10} & s_{11} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}
$$

The T-parameters are defined

$$
\begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{bmatrix} \begin{bmatrix} b_1 \\ a_1 \end{bmatrix}
$$

The relationship between S-parameters and T-parameters for a two-port network is,

$$
\begin{bmatrix} s_{00} & s_{01} \\ s_{10} & s_{11} \end{bmatrix} \implies \begin{pmatrix} -\Delta s & s_{00} \\ -s_{11} & 1 \end{pmatrix}
$$
$$
\Updownarrow \qquad\qquad \Updownarrow
$$
$$
\underbrace{\frac{1}{t_{11}} \begin{bmatrix} t_{01} & \Delta t \\ 1 & -t_{10} \end{bmatrix}}_{\text{S-parameters}} \Longleftarrow \underbrace{\begin{pmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{pmatrix}}_{\text{T-parameters}}
$$

Using T-parameters, cascading two networks ($\mathbf{X}$) and ($\mathbf{Y}$) together is calculated using matrix multiplication,

$$
\mathbf{X} \cdot \mathbf{Y} = \begin{pmatrix} -\Delta x & x_{00} \\ -x_{11} & 1 \end{pmatrix} \cdot \begin{pmatrix} -\Delta y & y_{00} \\ -y_{11} & 1 \end{pmatrix}
$$

De-embedding can be achieved by multiplying the matrix inverse of the T-parameters. This provides both a conceptually useful model of de-embedding, as well as a computationally efficient implementation. For example, to retrieve ($\mathbf{Y}$) from the cascaded connection of ($\mathbf{X} \cdot \mathbf{Y}$),

$$\mathbf{Y} = \mathbf{X^{-1}} \cdot (\mathbf{X} \cdot \mathbf{Y})$$

This is illustrated by Fig. 2.5.
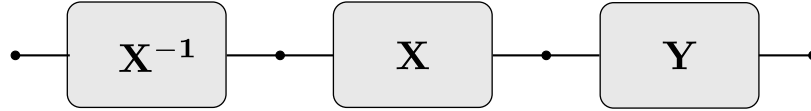
Figure 2.5.: Diagram illustrating cascading an inverse network

In general, the relationship between two-port networks can best be represented by a non-commutative algebra. Thus, matrix algebra is usually employed. However, a more powerful non-commutative algebra such as Geometric Algebra[21] could be used which may provide substantial benefits. A brief discussion about this is given in sec. 6.1.3.

Comments   The derivation of unterminating and de-embedding given here is in terms of reflection coefficient, but it can be equivalently formulated in terms of impedance or admittance. Penfield and Baurer's derivations in [14], for example, were made entirely in the context of impedance. Other than the different physical interpretations, this change amounts to a change of basis.

Although algebraically convenient, the mathematical implications of mixing independent and dependent variables in (2.2) are involved. One impact of this is to

complicate the statistical analysis of the residual errors. A comparison between a non-linear estimator and a linear one would make an interesting study.

2.2.2. Tiered Calibration

It is sometimes desirable to characterize an embedding network between some arbitrarily chosen pair of reference planes. This is accomplished by making enough measurements to unterminate at each plane, a technique referred to as a *tiered* calibration. A common reason for using a tiered calibration is to characterize a reciprocal, two-port device with a one-port system (see sec. 4.4.2 for a practical example).

A diagram illustrating a two-tiered calibration setup is shown in Fig. 2.6. In this example a pair of two-port networks, ($\mathbf{X}$) and ($\mathbf{Y}$) are connected in cascade, and network ($\mathbf{Y}$) is to be characterized. Assume that ($\mathbf{Y}$) cannot be measured in isolation, due to practical reasons. The procedure is to determine ($\mathbf{Y}$) from knowledge of ($\mathbf{X}$) and ($\mathbf{X} \cdot \mathbf{Y}$). This is accomplished by performing two calibrations at different *tiers*.
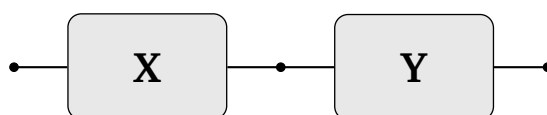
Figure 2.6.: Diagram of tiered embedding networks.

First Tier The first tier is created by removing ($\mathbf{Y}$), producing the circuit shown in Fig. 2.7 (this is the same scenario presented in Fig. 2.2). A sufficient number of calibration standards are attached at the y-pane, allowing ($\mathbf{X}$) to be determined.
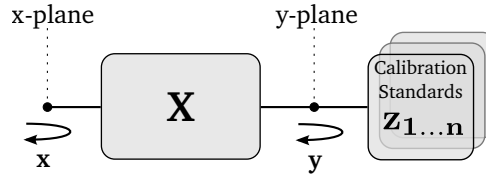
Figure 2.7.: Diagram of tiered embedding networks, illustrating the first tier.

Second Tier   Next, $(\mathbf{Y})$ is reconnected as in Fig. 2.8, and calibration standards are attached at the z-plane. This allows for the product of the two networks $(\mathbf{X} \cdot \mathbf{Y})$ to be determined. Combining the results from the first and second tiers, $(\mathbf{Y})$ can be found, by

$$\mathbf{Y} = \mathbf{X}^{-1} \cdot (\mathbf{X} \cdot \mathbf{Y})$$
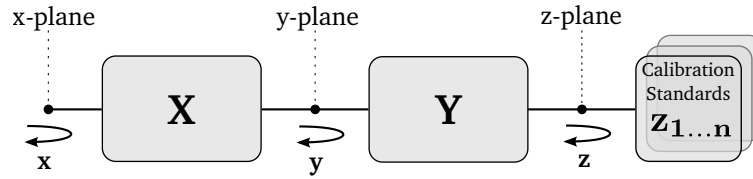
Figure 2.8.: Diagram of tiered embedding networks.

2.2.3. Extension to 2N-port VNA

Although the theory presented in the previous section was formulated in terms of one-port measurements, extension to two or more ports is fairly straight forward. This can be done most seamlessly by replacing the bilinear transform with the *matrical bilinear transform*, as described in [22]. Alternatively, the more popular formulation presented in [23] can be utilized.

Mathematically, the effect of increasing the number of ports is equivalent to increas-

ing the dimensionality of the vector space. An important subtlety of extending this analysis to two-port calibration is the inaccuracy of the *error-box* model, which has be presented here. This inaccuracy is due to the properties of an internal switch inside the VNA, and can be circumvented by accounting for the switch imperfections using what the community calls *switch-terms*[24]. Six port reflectometers, on the other hand, don't suffer from this discontinuity of representation.

2.2.4. The Calibration Processing Chain

As described above, the goal of calibration is to determine as accurately as possible the unknown parameters of the transformation, ($\mathbf{E}$). There are two possible reasons for this:

1. We want to remove the effects of ($\mathbf{E}$) from future measurements

2. We are interested in ($\mathbf{E}$) itself

Define the following:

$\mathbf{a}$	actual (true) reflection coefficient
$\mathbf{m}$	measured reflection coefficient
$\mathbf{i}$	ideal reflection coefficient of calibration standards
$\mathbf{c}$	corrected reflection coefficient
$\mathbf{E}$	true scattering parameters of embedding network
$\mathbf{E}'$	estimated scattering parameters of embedding network

The diagram shown in Fig. 2.9 illustrates the process of calibration. The actual reflection coefficient of a device ($\mathbf{a}$), is transformed by the embedding network ($\mathbf{E}$) to produce a measurement ($\mathbf{m}$). The known calibration standards, having ideal responses ($\mathbf{i}$), are measured in order to estimate the embedding parameters ($\mathbf{E}'$).

This is referred to as *unterminating.* It is necessary to differentiate the (**i**'s) from the (**a**'s) because the ideal response and actual response of a calibration standard are not identical. This difference is one cause of calibration error.
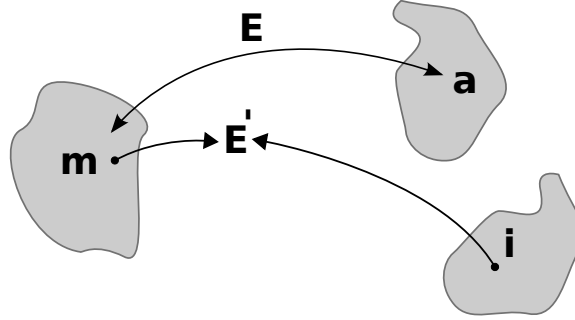
Figure 2.9.: The calibration processing chain illustrating the relationships between actual (**a**), measured(**m**), and ideal(**i**) responses.

A corrected response (**c**) is calculated by applying the inverse transform of the estimated parameters ($\mathbf{E}'$) to a measurement (**m**), this is *de-embedding.*

$$\mathbf{c} = f^{-1}\left(f\left(\mathbf{a}; \mathbf{E}\right); \mathbf{E}'\right)$$

The relation between the actual (**a**), measured(**m**), ideals(**i**), and corrected (**c**) responses is illustrated in Fig. 2.10. The calibrated response will, in practice, never be the identical to their actual responses because of unavoidable error in the estimation of the embedding network ($\mathbf{E}' \neq \mathbf{E}$).
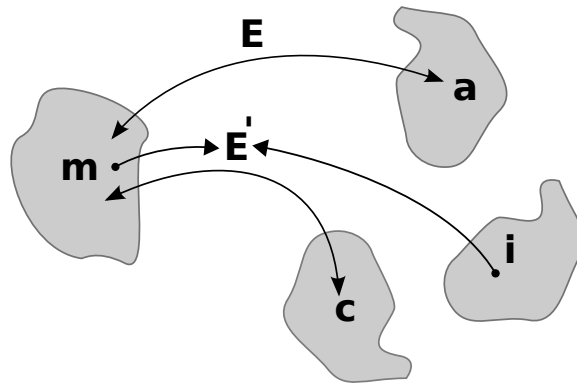
Figure 2.10.: The calibration processing chain illustrating the relationships between actual (**a**), measured(**m**), ideal(**i**), and corrected(**c**) responses.

It should be noted that the corrected measurements are shown to exist in a different space from both (**a**) and (**i**). This is done to represent the general case of an overdetermined calibration. In an overdetermined calibration, de-embedding the measurements of the calibration standards will not map them back to their definitions identically.

$$\mathbf{i} \neq f^{-1}\left(f\left(\mathbf{i}; \mathbf{E}\right); \mathbf{E}^{'}\right)$$

These differences are the residuals produced by the least squares estimator. The analysis of the residual errors is described in sec. 2.3.3.

Errors   To make things clear, various errors can be identified. There will always be some error in the calibration standard ideal definitions due to the inability to characterize the standards perfectly. At millimeter-wave and terahertz frequencies this error is significantly larger than at lower frequencies. This is mainly due to the machining precision and tolerances becoming significant compared to a wavelength.

The difference between the ideal response of a standard and the actual response of that standard is represented by (Δ_{ai}). This difference will cause create a difference

between the estimated embedding parameters ($\mathbf{E}'$) and the true embedding parameters ($\mathbf{E}$), written ($\Delta_{EE'}$), which in turn causes an error in the corrected response, (Δ_{ac}). Furthermore, all measurement errors are compounded by system noise, connector uncertainty, and other non-systematic errors, labeled (n). All of these errors, and how they relate to the calibration flow are illustrated in (Fig. 2.11).

While all of these errors exist, they are not available to the experimentalist. The only error that is perceivable is the lack-of-fit to the measurement model produced by an overdetermined calibration. This lack-of-fit is measured by the difference between the ideal responses and the corrected ideal responses ($\mathbf{\Delta_{ci}}$). These errors are the residuals from the least square estimator.
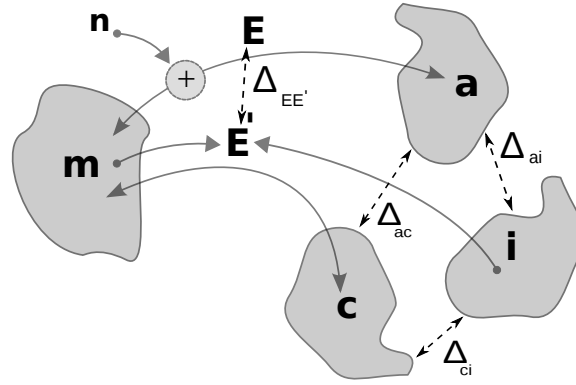
Figure 2.11.: The calibration processing chain with illustration of errors.

Variable	Difference between ...	Interpretation
Δ_{ai}	actual (true) reflection coefficient and ideal definition	error in assumed response of calibration standards
Δ_{ac}	actual (true) reflection coefficient and corrected response	corrected measurement error
Δ_{ci}	corrected reflection coefficient and ideal definition	residual error from least squares estimator
$\Delta_{EE'}$	actual (true) embedding network and estimated embedding network	error in embedding network estimation

Table 2.1.: Errors in calibration processing chain

2.2.5. Monte Carlo Simulation of Calibration

Due to the complexity of the calibration processing chain, the effects that different errors will have on the measurements of interest is not easily predicted. For example, having a phase error in a calibration standard definition will, in general, cause both phase and magnitude error in the corrected measurements. The way in which errors affect the measurements depends upon the embedding networks, which are different for each experiment. This makes diagnosing measurement problems or recommending appropriate tolerances difficult. To properly model errors in calibration, Monte-Carlo Simulations (MCS) must be used.

A MCS refers to a simulation in which all components of the calibration processing chain exist on a computer. Fictitious measurements, embedding networks, and error sources are all generated, allowing their effects on calibration error and measurement error to be analyzed. The basic flow of a MCS follows the diagram developed in Fig. 2.11, and is outlined below,

1. Create ideal definitions (**i**) and actual responses (**a**) of the calibration standards. Thus, error in the calibration standards (Δ_{ai}) is an input parameter.

2. Embed the actual responses in a realistic approximation of the embedding network (**E**), producing fictitious measurements (**m**). This embedding network is usually taken from an experiment, i.e. a previous calibration made on the system of interest.

3. Optionally, add noise and other errors sources to the measurement (electronic drift, connection error, etc.), either before or after (**E**) as appropriate. For example, connector misalignment effects (**a**), while electronic noise is added to (**m**).

4. Using the ideal definitions (**i**) and their measurements (**m**), run a calibration,

which produces an estimate of the embedding network ($\mathbf{E}'$). At this point the residual error (Δ_{ci}) and estimated embedding network error ($\Delta_{\mathbf{EE}'}$) are both available.

5. Create fictitious measurements of DUT's following steps 2-3 and calculate corrected measurements by de-embedding the estimated embedding network ($\mathbf{E}'$). This provides the measurement error (Δ_{ac}).

Once the MCS has completed, all of the errors in Tab. 2.1 may be calculated. Numerous trials can be run and the resultant errors can be statistically analyzed to produce confidence bounds on the quantities of interest. Examples of synthetic experiments are given in sec. 3.5.2, sec. A.2, and sec. E.3.

2.3. Calibration Uncertainty

The experimentalist must have a technique to estimate uncertainty in their measurements so that results can be asserted as *true* with some amount of confidence. Quantitative uncertainty estimation prevents unwarranted conclusions being drawn from experimental data, and precludes the unnecessary modeling of measurement artifacts. This section introduces some known causes of uncertainty in VNA measurements at terahertz frequencies, followed by a description of the mathematical effects of such uncertainties on calibration.

2.3.1. Sources of Uncertainty

VNA noise

Any microwave source or detector system produces noise. The VNA frequency extension heads increase the frequency of the internal VNA source by using a series of

multiplier chains that amplify existing, and add additional, noise. The VNA electronic noise sets the *noise floor*, below which, nothing meaningful can be measured. Thus, it is an important quantity to measure before attempts to estimate other amounts of uncertainty, such as connector repeatability, are made.

To qualitatively illustrate typical VNA noise at 500-750GHz (WR-1.5/WM-380), a plot of the complex drift of uncalibrated data is shown in Figure Fig. 2.12. This is a plot of the uncalibrated complex reflection coefficient for three different standards measured continuously over 3hrs, producing 180 sweeps. Each swirled data cluster is the reflection coefficient at a single frequency, over the 180 sweeps.
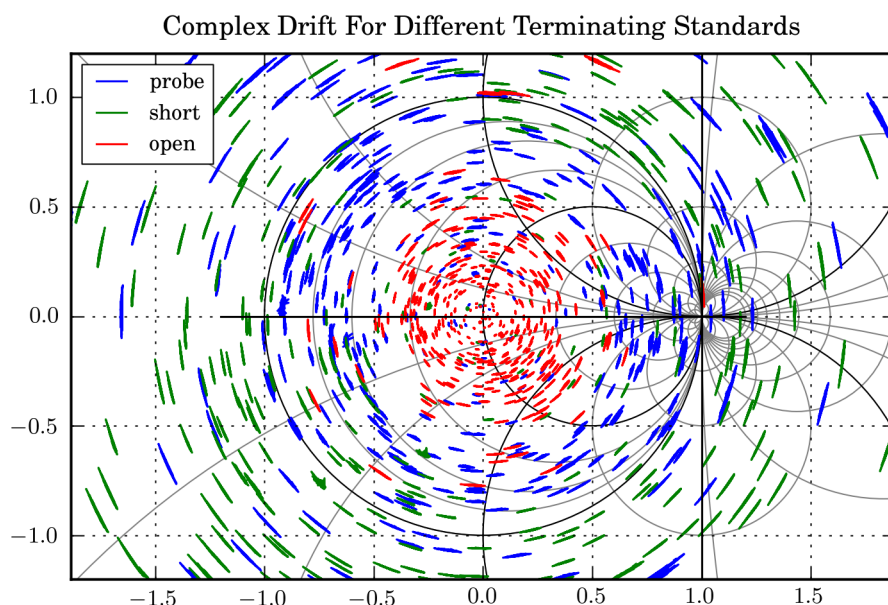
Figure 2.12.: Visualization of Drift: Uncalibrated Reflection Measurements Taken Over Time

This qualitative illustration is a useful starting point for the analysis. Firstly. it provides a coarse estimate for the precision of the system, but more importantly, it indicates that the noise is most accurately described as *curling.* This allows two important conclusions to be drawn: 1) the noise is more naturally decomposed in

polar components as opposed to cartesian and 2) the uncertainty in the phase component is much larger than in the magnitude component. In other words, the phase noise is dominant. Effectively, these casual observations form a principle component analysis of the random vector noise signal, achieved solely through qualitative methods.

Connector Repeatability

Another important source of error in VNA calibrations is the random effects caused by connector repeatability[25]. In this context *connector* repeatability is interpreted in an abstract sense, and has different interpretations in different media. In rectangular waveguide it represents the flange connection, while in quasi-optical measurements it may be the replacement of a device in a fixture. Recently, a method was presented that can be used to determine a model for the physical sources of coaxial connector uncertainty [26]. This approach uses the frequency dependence of calibration standards' variance that may be estimated by redundant measurements. This method can be extended to different media, provided a reasonably accurate model of the physical cause of uncertainty is available.

Cable Drift

Cable drift (which is a misleading term) is error induced by the change in the electrical properties of the cabling that connects the frequency extenders to the VNA. It is one of the dominant sources of uncertainty in two-port measurements due to the fact that the reference and measurement signals are disturbed independently. Because most of the work in this thesis concentrates on one port measurements, this source of uncertainty was not investigated in detail. It is mentioned here because it is

an important, and perhaps dominant, source of uncertainty for two-port applications at terahertz frequencies.

2.3.2. Biased vs. Unbiased Errors

Some sources of measurement uncertainty have been discussed, and now the effects of such uncertainties from a mathematical perspective will be explored. Generally speaking, when measurements are fit to a model there will be discrepancies. These descrepencies are usually called *errors* (or residual errors) and are classified into two categories: **biased** and **unbiased**. Biased errors effect measurement **accuracy**, while unbiased errors effect measurement **precision**. The amount of unbiased error (precision) can be estimated by redundant measurements. Biased errors (accuracy) can, in some cases, be estimated through over-determined measurements. A table summarizing these relationships is shown in Table Tab. 2.2, accompanied by a illustration in Fig. 2.13.

Error Type	Effects Measurement ...	Estimated by ...
Biased	**Accuracy**	**Overdetermined Measurements**
Unbiased	**Precision**	**Redundant Measurements**

Table 2.2.: Errors: Their Effects and Estimation Techniques
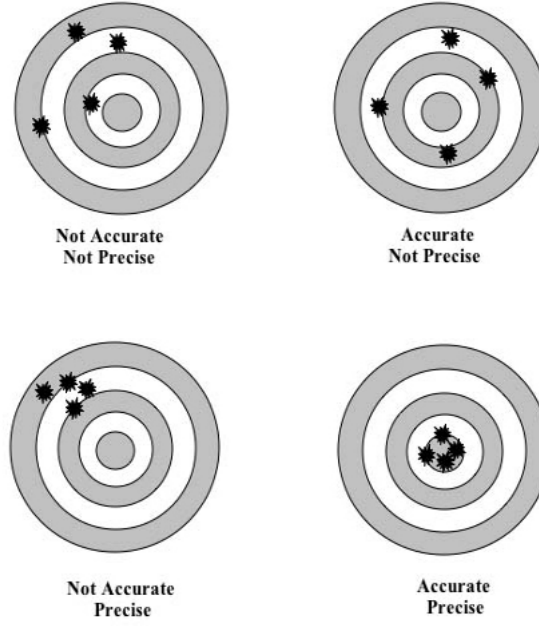
Figure 2.13.: Illustration of precision and accuracy (taken from noaa.gov)

Given an overdetermined, redundantly measured set of calibration standards, the residuals can be analyzed to estimate of the amount of biased and unbiased errors. Most discussions of complex residual analysis approach the problem from a analytical perspective[27, 28, 29], drawing on concepts from statistics. In contrast to these approaches, the following residual analysis is based solely on geometrical arguments. This geometrical arguments are then used to create quantitative metrics of estimate biased and unbiased errors. Although lacking in mathematical rigor, the error metrics presented here have been used thoroughly in practice and proven to be useful. Examples are presented in sec. A.5.

2.3.3. Residual Analysis

Given a set of n measurement/ideal pairs $(\mathbf{m_{1\dots n}}, \mathbf{i_{1\dots n}})$, we can calculate an estimate for the embedding network $(\mathbf{E}')$. This embedding network estimate can be used to

calculate corrected responses for those same measurements ($\mathbf{i}'$),

$$
\begin{aligned}
\mathbf{i}' &= \frac{e_{00} - \mathbf{m}}{\Delta_e + \mathbf{m}e_{11}} \\
&= \mathbf{E}'^{-1} \circ \mathbf{m}
\end{aligned}
$$

Where the inverse ($\mathbf{E}'^{-1}$) represents inverse T-parameters, and open dot ($\circ$) implies the cascading operation. If the set is overdetermined, meaning more than 3 measurement-ideal pairs are available for a one-port calibration, then there will be a general lack of fit between the given $\mathbf{i}$'s and the corrected $\mathbf{i}'$'s. A plot of a typical set of calibration standard ideals, and the corrected measurements of the ideals is shown below in Fig. 2.14.
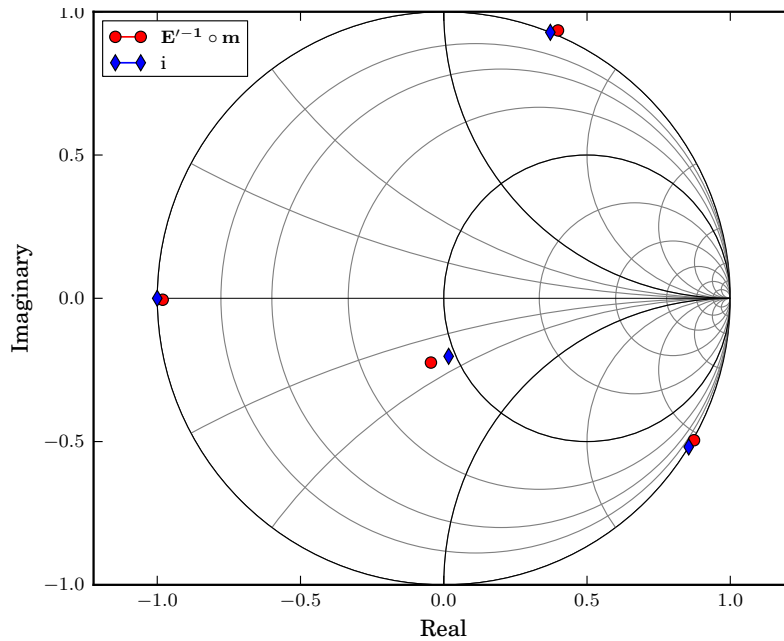
Figure 2.14.: Corrected calibration standards and their ideal response for an overdetermined calibration.

For each measurement, this lack of fit produces a residual which is the complex difference between the de-embedded measurement and ideal response,

$$
\begin{aligned}
\delta &= \mathbf{i} - \mathbf{i}' \\
&= \mathbf{i} - \mathbf{E}'^{-1} \circ \mathbf{m}
\end{aligned}
\tag{2.7}
$$

It is noted that the analysis could take place in the m-plane, by calculating the residuals using,

$$
\begin{aligned}
\delta &= \mathbf{m} - \mathbf{m}' \\
&= \mathbf{m} - \mathbf{E}' \circ \mathbf{i}
\end{aligned}
$$

However, there are many disadvantages in doing this. The largest problem is that the units of the residuals are determined by the properties of the embedding network ($\mathbf{E}$), which makes interpretation less valuable. All of the residuals calculated in this work are done using (2.7).

By constructing a redundantly measured, over-determined calibration a set of complex residuals, defined by (2.7), is produced. A typical plot of complex residuals is shown in Fig. 2.15. These residuals were generated from a one-port calibration consisting of four standards, each measured eight times. The sets of identical markers correspond to measurements of a single standard.
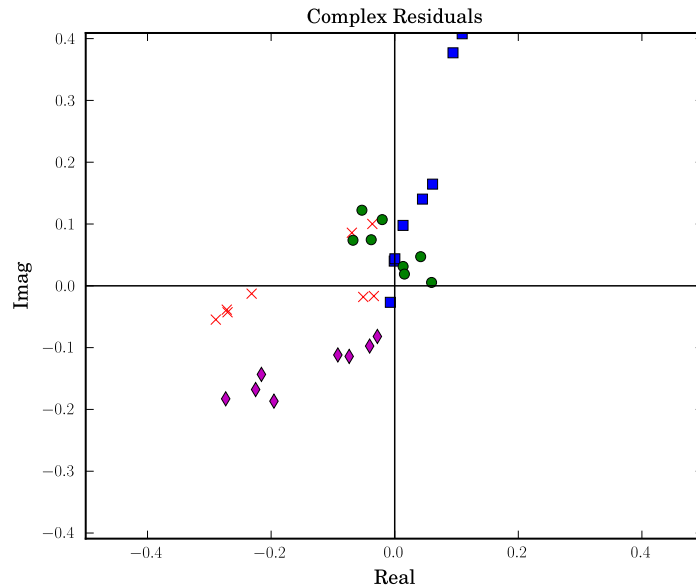
Figure 2.15.: Complex residual errors for an overdetermined, redundantly measured set of standards.

Using a linear least-squares estimator in the presence of random, white Gaussian noise produces residuals that form a zero-mean normal distribution [30]. From the plot in Fig. 2.15, it can be seen the groups of errors from each standard do not have a zero mean. A non-zero mean indicates that there is a lack-of-fit between the ideal responses and their de-embedded measurements. It is this observation which forms the basis for the biased vs unbiased error metrics in the next section.

2.3.4. Biased and Unbiased Error Metrics

In this section metrics are formulated to estimate biased and unbiased errors from the complex residuals produced by a least squares estimator. These quantities are defined heuristically, lacking a rigorous mathematical basis, but have been used in practice and proven to be a valuable tool. Shown in figure Fig. 2.16 is an illustration of a typical set of residuals.
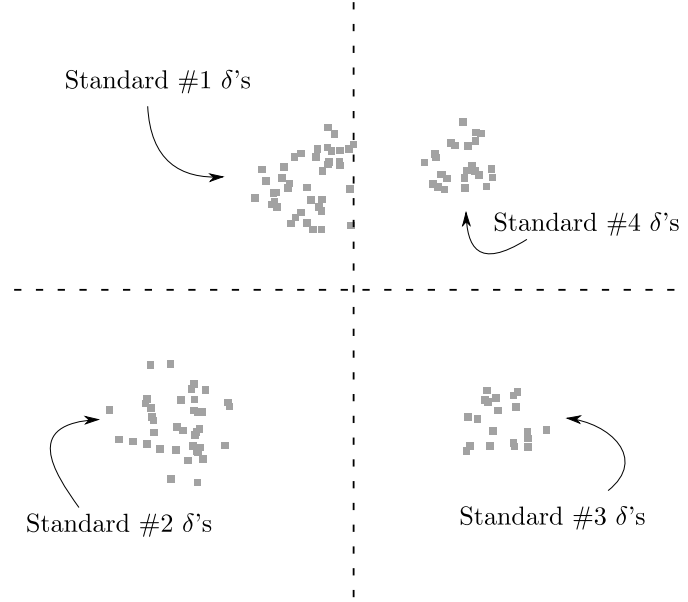
Figure 2.16.: Illustration of the distribution of residuals in the presence of biased and unbiased error.

For each measurement there is a corresponding complex residual (δ), represented by a small square marker in Fig. 2.17. All of the residuals for a single calibration standard form a complex set or *cloud*, having a mean (μ) and a standard deviation (σ). Given by the familiar formulas,

$$
\begin{aligned}
\mu &= \frac{1}{N} \sum_{n}^{N} \delta_n \\
\sigma &= \sqrt{\frac{1}{N} \sum_{n}^{N} |\delta_n - \mu|^2}
\end{aligned}
\tag{2.8}
$$

Where N is the number of measurements per standard. The diagram in figure Fig. 2.17 is re-drawn in Fig. 2.17 with labels indicating the these quantities.
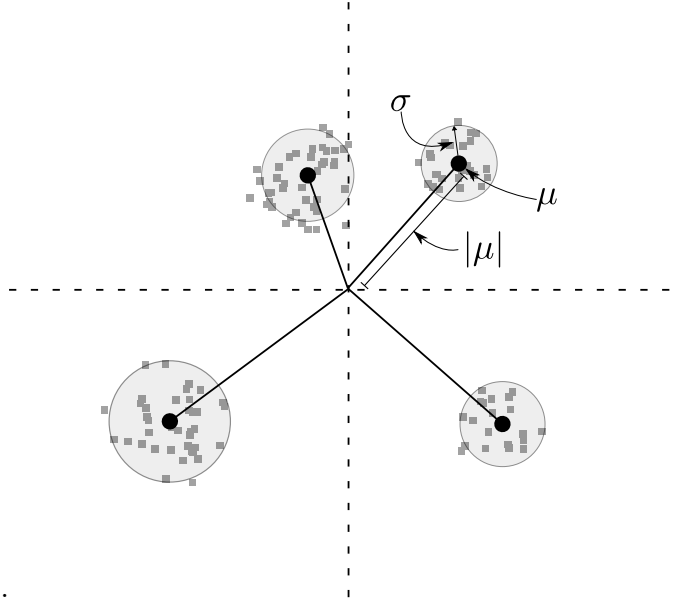
Figure 2.17.: Illustration of distribution of residuals in the presence of biased and unbiased error with statistical quantities labeled.

From Fig. 2.17, it is possible to create metrics to measure both biased and unbiased errors, independently. The biased error manifests itself by displacing the mean of a single standard's residual cloud. The unbiased error produces a larger spread in the residual cloud. Defining the following metrics:

- **unbiased error:** Average spread of each residual cloud.

$$\frac{1}{M} \sum_m^M \left(\sigma_m \left(\delta_n \right) \right)$$

  where (M) is the number of standards, and (N) is the number of measuarments.

- **biased error:** Average distance of each cloud's center of mass from the origin.

$$\frac{1}{M} \sum_m^M \left| \mu_m \left(\delta_n \right) \right|$$

- **total error:** Average magnitude of all residuals.

$$\frac{1}{N} \sum_{n}^{N} |\delta_n|$$

A plot of these metrics vs. frequency for a calibration made at 325-500GHz rectangular waveguide (WR-2.2) is shown below in Fig. 2.18. Plotting these metrics on a log scale is convenient because, it allows the metrics to be interpreted as the smallest measurable return loss of the system. For example, Fig. 2.18 is indicates that the calibration accuracy is limited by biased error to at least -30 and -20dB. This is most likely due to improperly defined calibration standards. The system precision, however, is approximately -40dB, which indicates there is a potential for improvement if the cause of the systematic error can be identified and removed. This plot was taken from the measurement study presented in sec. A.5.
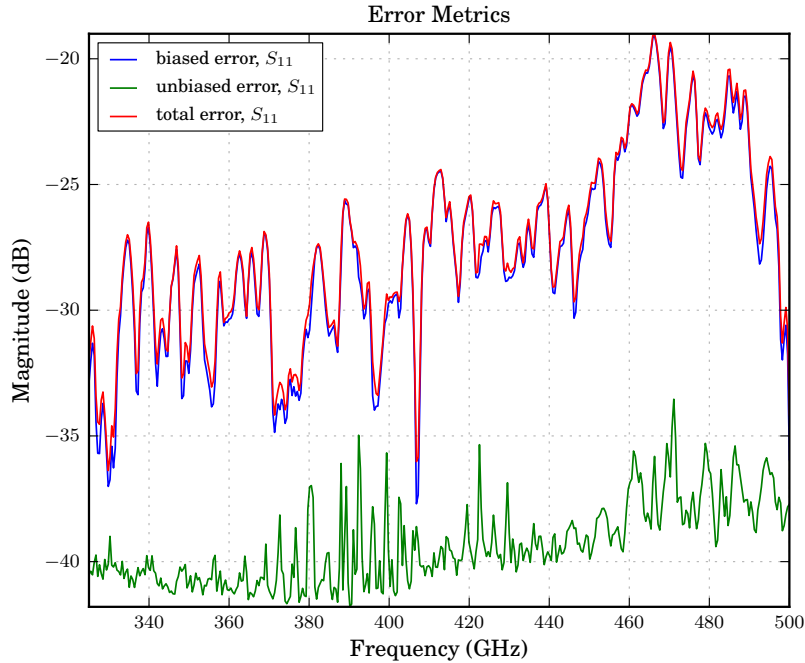
Figure 2.18.: Example application of the residual error metrics for an overdetermined redundant calibration at 325-500GHz

Multi-frequency Residuals In practice, redundant measurements are not always made by the experimenter, so the above analysis is not directly applicable. However, because measurements are available at a number of frequencies, the biased error can still be estimated. Shown in Fig. 2.19 are the complex residuals taken from an overdetermined calibration having systematic errors in the delay short standard lengths. The systematic error manifests itself as correlation between residuals at different frequencies. To apply the metrics defined above, the statistical functions defined in (2.8) are calculated over frequency, as apposed to over redundant measurements. This will create a single number for each of the metrics defined above, as opposed to traces across frequency. Regardless, the metrics are interpreted identically.
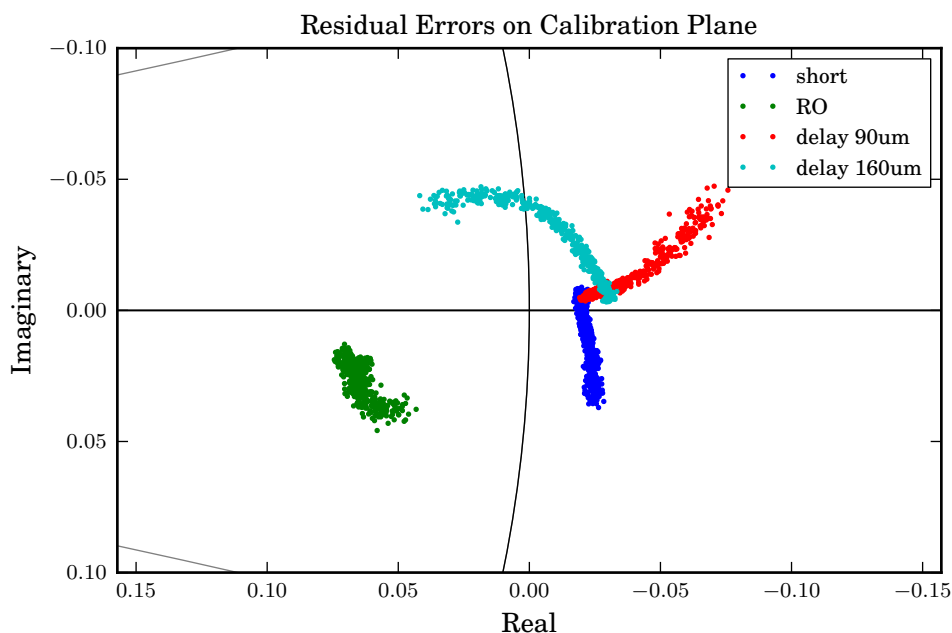
Figure 2.19.: Complex residual errors for single re-connection of calibration standards, but for many frequencies

Measuring more than the minimal number of standards provides a wealth of information about the calibration. This information can be used to estimate the accuracy

of a calibration and aid in the diagnosis of any errors that are present. Such capabilities are extremely useful for non-conventional and experimental measurements, in which systematic errors are common.

The next section presents a brief discussion of the Möbius transformation, a general class of transformations that encompasses the embedding relationship of 2.1. Some well known properties of the Möbius transformation are interpreted in the context of unterminating and de-embedding, which may be of use to the microwave engineer. None of the following chapters directly depend on the contents of this section, and therefore it may be skipped without consequence.

2.4. *Möbius Transformation

2.4.1. Introduction

It is worthwhile studying the Möbius transformation (also known as the bi-linear transformation or fractional linear transformation) because it occurs frequently in the art of microwave engineering. For example, the relationship between impedance and reflection coefficient, as well as the action of cascading networks are both Möbius transformations. This transform appears in other interesting fields as well such as the theory of relativity [31, 32] and computer graphics[33]. The study of Möbius transformation is not merely a pedagogical exercise. Rather, it integrates and unifies developments in seemingly dissimilar fields.

The Möbius transformation is defined as a rational function of the form[32],

$$\mathbf{w} = f\left(\mathbf{z}; a, b, c, d\right) = \frac{a\,\mathbf{z} + b}{c\,\mathbf{z} + d} \tag{2.9}$$

$$ad - bc \neq 0$$

Where, $\mathbf{w}$ and $\mathbf{z}$ are the complex variables being transformed, and $a, b, c,$ and d, are complex constants which are the parameters of the transform. Its inverse exists and is defined,

$$\mathbf{z} = f(\mathbf{w}; a, b, c, d) = \frac{d\,\mathbf{w} - b}{-c\,\mathbf{w} + a} \tag{2.10}$$

$$ad - bc \neq 0$$

The cascading relationship can be put into this standard form which allows the properties of the Möbius transformation to be interpreted in the context of calibration. Starting from the measurement equation relating the measured reflection coefficient ($\mathbf{m}$), to the actual reflection coefficient ($\mathbf{a}$) and embedded network ($\mathbf{E}$),

$$\mathbf{m} = f(\mathbf{a}; E) = e_{00} + \frac{e_{01}e_{10}\,\mathbf{a}}{1 - e_{11}\,\mathbf{a}} \tag{2.11}$$

Treating the determinant of the embedding network's s-parameter matrix as a single variable (Δe), then (2.11) can be manipulated into the standard form.

$$\mathbf{m} = \frac{-\Delta e\,\mathbf{a} + e_{00}}{-e_{11}\,\mathbf{a} + 1} \tag{2.12}$$

Applying the result that the inverse of the Möbius is given by (2.10), we can immediately write down the formula for the actual reflection coefficient in terms of the

measured reflection coefficient and embedding s-parameters.

$$\mathbf{a} = \frac{\mathbf{m} - e_{00}}{e_{11}\,\mathbf{m} - \Delta e}$$

This is the formula used for de-embedding.

2.4.2. Decomposition Into Simpler Functions

The cascading relationship can be decomposed into a series of simpler functions, each dependent upon one parameter of the embedding network. This decomposition allows one to gain some intuition about how the parameters of the embedding network transform the reflection coefficient space, which proves to be useful in calibration problems.

Starting from the measurement equation ,

$$\mathbf{m} = f\left(\mathbf{a}\right) \quad = \quad e_{00} + \frac{e_{01}e_{10}\cdot\mathbf{a}}{1 - e_{11}\cdot\mathbf{a}} \tag{2.13}$$

Dividing the second term by $\mathbf{a}$, we have

$$\mathbf{m} \quad = \quad e_{00} + e_{01}e_{10}\cdot\frac{1}{\frac{1}{\mathbf{a}} - e_{11}} \tag{2.14}$$

In this form, (2.14) can clearly be decomposed into a a series of simpler functions,

$$\mathbf{m} = f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1(\mathbf{a})$$

$$f_1(z) = 1/z \qquad\qquad \text{inversion}$$

$$f_2(z) = z - e_{11} \qquad\qquad \text{translation}$$

$$f_3(z) = 1/z \qquad\qquad \text{inversion}$$

$$f_4(z) = e_{10}e_{01} \cdot z \qquad\qquad \text{dilation/rotation}$$

$$f_5(z) = z + e_{00} \qquad\qquad \text{translation} \qquad\qquad (2.15)$$

The result of the individual functions acting successively on a passive polar grid ($|z| \leq 1$) is illustrated in Figure (Fig. 2.20). This can be thought of as demonstration of an embedding network on the region of passive devices.
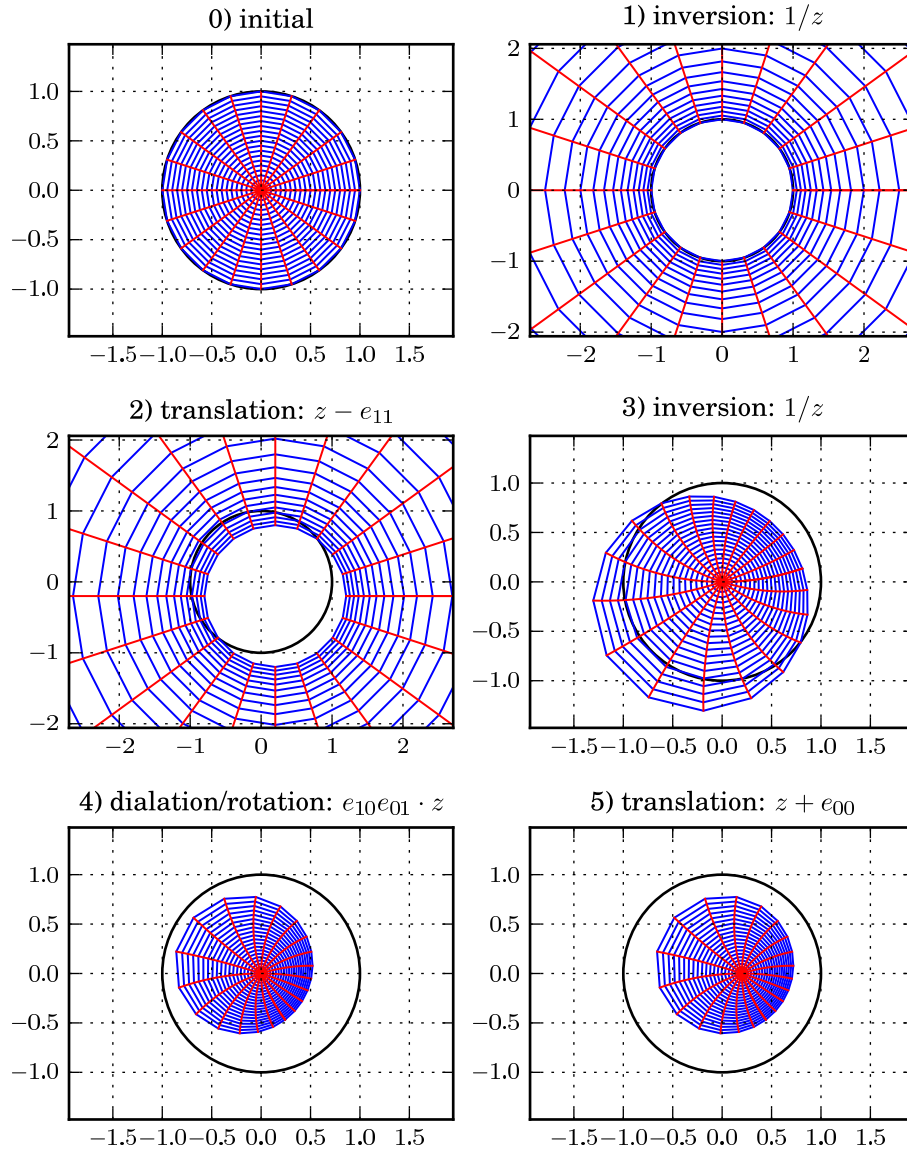
Figure 2.20.: Illustration of the embedding equation broken down into individual sub-transforms, each dependent upon a single s-parameter of the embedding network.

From this figure some specific effects can be attributed to circuit embedding parameters. The first three sub-functions $(f_1 \cdots f_3)$ invert the plane in the unit circle, translate by e_{11}, and then invert again. This specific combination of transforms is

known as a *special conformal transformation*[32]. It is the feedback created by the e_{11} parameter that is responsible for the conformal warping of the space. In such a warping, circles are preserved, although their centers are not.

The remaining sub-functions make up an affine transformation. The effect of the combined $e_{01}e_{10}$ term is a simple linear rotation/dilation, followed by a translation by e_{00}.

2.4.3. Projective Geometry

The fact that the Möbius transformation can be implemented by matrix multiplication is a result of projective geometry[34, 35, 32]. While a full account of projective geometry is beyond the scope of this dissertation, the general concept can be explained, and the results are simple to follow.

The underlying idea is to represent a vector in some original space, by a vector in a higher dimension space, and relate the two spaces by a projection operation. An abstract example of this relationship is illustrated in Fig. 2.21. In this figure, a vector in three dimensions is related to vector in two by where it intersects some some projection plane. Given this construction, a linear transformation acting on the higher dimensional space turns out to be a Möbius transformation in the lower dimensional one. Thus, one exchanges a complicated transformation in a simple space, for a simple transformation in a more complicated space. This is analogous to using the Fourier transformation to linearize convolution.
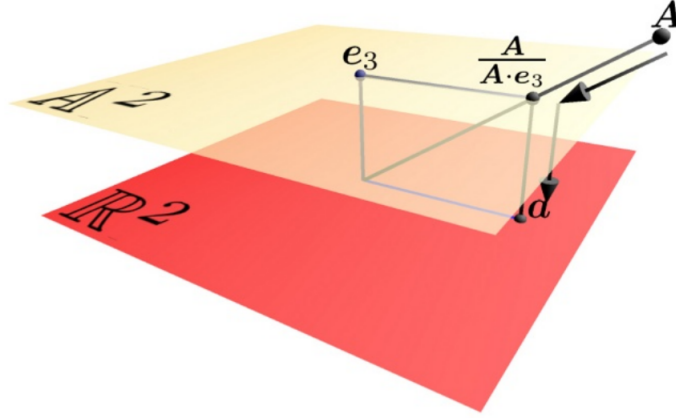
Figure 2.21.: Illustration of projective geometry. The three dimensional vector (**A**) is projectively related to the two dimensional vector (a). Taken from [33]

Projective Coordinates   Through the use of projective (homogeneous) coordinates[36, 32] the Möbius transformation can be realized as matrix multiplication. A projective coordinate is simply representing a single number by the division of a pair of numbers.

$$\mathbf{z} = \frac{z_1}{z_2} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \tag{2.16}$$

If one thinks of z_1 and z_2 as being cartesian components of a two dimensional vector, then **z** represents the direction of the vector. Using soft brackets for projective coordinates to imply division of the first and second rows. Given this representation, the Möbius transformation can be written as a 2x2 matrix,

$$\mathbf{w} = \frac{a\,\mathbf{z} + b}{c\,\mathbf{z} + d}$$

$$\begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix} nd \tag{2.17}$$

Thus the Möbius transform is a linear transform of the projective coordinates, and can now be implemented by matrix multiplication. The subsequent operation of two Möbius transformations (f_1) and (f_2), is

$$\mathbf{w} = f_2(f_1(\mathbf{z}))$$

$$f_2 \circ f_1 \circ \mathbf{z} = F_2 \cdot F_1 \cdot \mathbf{z}$$

$$\frac{a_2 \left(\frac{a_1\,\mathbf{z}+b_1}{c_1\,\mathbf{z}+d_1} \right) + b_2}{c_2 \left(\frac{a_1\,\mathbf{z}+b_1}{c_1\,\mathbf{z}+d_1} \right) + d_2} = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix}$$

$$\frac{(a_2 a_1 + b_2 c_1)\,\mathbf{z} + (a_2 b_1 + b_2 d_1)}{(c_2 a_1 + c_1 d_2)\,\mathbf{z} + (c_2 b_1 + d_2 d_1)} = \begin{pmatrix} a_2 a_1 + b_2 c_1 & a_2 b_1 + b_2 d_1 \\ c_2 a_1 + c_1 d_2 & c_2 b_1 + d_2 d_1 \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix}$$

Stringing together a series of Möbius transformations is reduced to a series of matrix multiplications. This provides both conceptual and computational advantages. For example, applying the simple result from linear algebra that the inverse of a square matrix exists provided the determinant is not zero, allows us to calculate the inverse Möbius transform.

$$\begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix} \tag{2.18}$$

$$\begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ 1 \end{pmatrix} \tag{2.19}$$

$$\mathbf{z} = \frac{d\,\mathbf{w} + -b}{-c\,w + a} \tag{2.20}$$

Cascading Networks Applying these ideas to the cascading relation:

$$\mathbf{m} = \frac{-\Delta e \cdot \mathbf{a} + e_{00}}{-e_{11} \cdot \mathbf{a} + 1}$$

$$\begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} = \begin{pmatrix} -\Delta e & e_{00} \\ -e_{11} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ 1 \end{pmatrix} \tag{2.21}$$

From (2.18), the inverse can be immediately written down,

$$\begin{pmatrix} \mathbf{a} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & -e_{00} \\ e_{11} & -\Delta e \end{pmatrix} \begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} \tag{2.22}$$

This is the formula used for de-embedding. This projective matrix representation of a two-port networks is known by the microwave community as *scattering transfer parameters*, or T-parameters[20]. As stated above, this allows many two-port networks can be cascaded together using matrix multiplication. For example, the network depicted in Fig. 2.8 has the matrix representation.

$$\begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} = \mathbf{X} \cdot \mathbf{Y} \cdot \begin{pmatrix} \mathbf{a} \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} -\Delta x & x_{00} \\ -x_{11} & 1 \end{pmatrix} \begin{pmatrix} -\Delta y & y_{00} \\ -y_{11} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ 1 \end{pmatrix}$$

The decomposition of the transformation as discussed in sec. 2.4.2, (2.21) can be broken apart according to (2.15) like so

$$\mathbf{m} = f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1 (\mathbf{a})$$

$$\begin{pmatrix} \mathbf{m} \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & e_{00} \\ 0 & 1 \end{pmatrix}}_{f_5} \overset{\text{translation}}{} \underbrace{\begin{pmatrix} e_{10}e_{01} & 0 \\ 0 & 1 \end{pmatrix}}_{f_4} \overset{\text{rot/scale}}{} \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{f_3} \overset{\text{inversion}}{} \underbrace{\begin{pmatrix} 1 & -e_{11} \\ 0 & 1 \end{pmatrix}}_{f_2} \overset{\text{translation}}{} \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{f_1} \overset{\text{inversion}}{} \begin{pmatrix} \mathbf{a} \\ 1 \end{pmatrix}$$

Impedance Reflection Coefficient Applying the matrix representation to the impedance reflection coefficient relation.

$$\Gamma = \frac{\mathbf{Z_l} - Z_0}{\mathbf{Z_l} + Z_0}$$

If one normalizes the impedance ($\mathbf{z} = \frac{Z_l}{Z_0}$) and replaces (Γ) with ($\mathbf{s}$) for consistency of notation, this becomes,

$$\mathbf{s} = \frac{\mathbf{z} - 1}{\mathbf{z} + 1}$$

$$\begin{pmatrix} \mathbf{s} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix}$$

Which has inverse,

$$\begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{s} \\ 1 \end{pmatrix}$$

The relation between normalized impedance ($\mathbf{z}$) and normalized admittance ($\mathbf{y}$) can is expressed,

$$\mathbf{z} = \frac{1}{\mathbf{y}}$$

$$\begin{pmatrix} \mathbf{z} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ 1 \end{pmatrix}$$

From these results all relations between $\mathbf{z}, \mathbf{y},$ and $\mathbf{s}$ can be derived quickly.

2.5. Conclusion

The fundamentals of calibration have been presented in the more general context of unterminating and de-embedding. The calibration processing chain has been described in detail, and various errors within it have been identified. These errors led to a geometrically-based complex residual error analysis, which has been used to create practical metrics of biased and unbiased errors. The following chapters apply the ideas and techniques developed in this chapter to practical calibration and measurement problems in the field of terahertz metrology.

3. Multi-Frequency One-Port Self-Calibration using Parametric Standards

This chapter presents a general approach to multi-frequency one-port self-calibration using parametric standards. The approach given here determines optimal parameters for partially defined calibration standards through the minimization of residual error. As a demonstration of the method, an algorithm is developed which uses an iterative least squares solution for parameter estimation. The algorithm is applied to a calibration set containing delay shorts of unknown physical length, and then refined to model a delayed load of unknown length and reflection coefficient magnitude. Measurements are performed at 500-750GHz (WR-1.5, WM-380) and show that the self-calibration provides improved measurement accuracy.

3.1. Introduction

As waveguide and on-chip metrology continues to be pushed toward the Terahertz spectrum accurate characterization of calibration standards becomes increasingly difficult. This difficulty is due to two main reasons: dimensional uncertainty, and connection repeatability. Although efforts to solve these problems through improved manufacturing accuracy and better hardware design are justified[9, 2], much can be done to address the problem analytically[37].

As a solution to this problem, the application of multi-frequency, one-port self-calibration using parametric standards is investigated. We present a general theoretical foundation (Section sec. 3.3), followed by a specific implementation utilizing an iterative least squares approach (Section sec. 3.4). This implementation is then applied to practical examples of rectangular waveguide calibration at 500-750GHz using delayed shorts and delayed loads of unknown lengths (Section sec. 3.5.). The multi-frequency self-calibration technique shows improvement over conventional one-port calibration, given the same initial standard definitions.

All algorithms and calculations presented in this work have been made with the open-source python module scikit-rf[38].

3.2. Previous Work

The general idea of improving the accuracy of de-embedding and unterminating by using redundant measurements was first presented in [14]. Elaborating upon this concept, modern calibration procedures make precise use of additional measurement information to solve for specific unknowns in the calibration set itself, a technique known as self-calibration [39]. Such self-calibration algorithms can be categorized

as closed-form [23, 40, 41, 22] or iterative[28, 42, 29]. While elegant and fast, closed-form self-calibration methods are only applicable to the specific calibration kit for which they were designed. Iterative methods, on the other hand, provide the benefits of self-calibration without the limitation of a single, preconceived calibration set. Thus, a single algorithm can handle a variety of calibration sets, allowing for additional and redundant measurements to be added easily. The drawback of iterative methods is that the existence of a unique solution is not guaranteed, and the computation speed is slower.

While much focus has been spent on two-port self-calibration schemes, related developments for one-port calibrations are still of interest. Attempts to make accurate measurements at millimeter wave and sub-millimeter wave frequencies, where precise line and load standards are difficult to realize, has produced a need for more elaborate one-port algorithms [37]. In response, several closed-form one-port self-calibration algorithms have been developed which utilize partially known standards. One such solution uses two reflective standards of unknown phase [43]. This solution was further extended in [7] by replacing the matched load standard with an arbitrary load. The use of radiating open[12] as a load standard was then used to produce a calibration procedure resistant to flange misalignment.

Because there is less information provided by a one-port measurement, the potential applications of closed-form self-calibration methods are limited. One solution to this problem, originally presented in [44], is to exploit the inter-frequency relationship of the calibration standards in order to reduce the number of unknowns in the calibration kit itself. The concept originates from the observation that the unknown parameters in a calibration kit may be physical quantities not changing with frequency. For example, one may know the propagation constant of the medium precisely, but not the physical length of the delay lines. Or perhaps one may have a

closed form expression for a connector discontinuity as a function of some unknown dimensional quantity. This technique, referred to as multi-frequency calibration (MFC), has been applied in recent years to both one-port and two-port calibration [45, 37, 44, 46] .

The work presented here provides a simple approach to iterative self-calibration based on the minimization of residual error. In contrast to the work in [37, 44, 46], our theoretical foundation is sufficiently general to allow for application to any calibration kit. An equivalently general approach to MFC has been independently developed in [45]. While the work in [45] is clearly written and conceptually useful, the details provided are insufficient for implementation, and in-depth measurement results are absent. The work here is unique in its simplicity and transparency, thereby lowering the barrier for others to make use of the technique. Furthermore, we present practical application to a relevant measurement challenge that indicates the effectiveness of the procedure. .

3.3. Theory

A raw measurement ($\mathbf{m}$) is a function of the device's actual response at the reference plane ($\mathbf{a}$) as well as the effects of the embedding circuitry of the measurement system (E).

$$\mathbf{m} = f\left(\mathbf{a}; E\right) \tag{3.1}$$

Where E is some set of complex parameters, such as $E = \{e_{00}, e_{11}, e_{01}e_{10}\}$ in the case of one-port calibration. The goal of calibration is to measure a sufficient set

of known responses ($\mathbf{a}_{1\ldots i}$), so that the effects of the embedding circuitry can be removed from future measurements. Given that the embedding circuitry can be solved for, (3.1) can be inverted to calculate a corrected response ($\mathbf{c}$) from the raw measurement ($\mathbf{m}$) of a device under test.

$$\mathbf{c} = f^{-1}\left(\mathbf{m}; E\right) \tag{3.2}$$

Iterative approaches to calibration determine the unknown embedding parameters by minimizing the mean square difference between the ideal responses of the calibration standards, and their de-embedded responses.

$$\epsilon^2 = \sum_{i=1}^{I} \left| \mathbf{a}_i - f^{-1}\left(\mathbf{m}_i; E\right) \right|^2 \tag{3.3}$$

$$min(\epsilon^2; E) \tag{3.4}$$

Where ϵ is the residual error, and I is the number of calibration standards. Self-calibration methods allow for parameters of the calibration standards ($\mathbf{a}_{1\ldots i}$) to be solved for during the determination of the embedding circuitry (E). Thus, a single response ($\mathbf{a}_i$) is a function of some unknown, to-be-determined set of p-parameters ($\varphi_{1\ldots p}$).

$$\mathbf{a}_i\left(\varphi\right) = f\left(\varphi_0, \varphi_1, \ldots \varphi_p\right) \tag{3.5}$$

The residual error then becomes,

$$\epsilon^2 = \sum_{i=1}^{I} \left| \mathbf{a}_i\left(\varphi\right) - f^{-1}\left(\mathbf{m}_i; E\right) \right|^2 \tag{3.6}$$

The adjustable parameters of the minimization is extended to all parameters, φ .

$$min(\epsilon^2; E, \varphi) \tag{3.7}$$

Multi-frequency calibration methods allow for a single unknown parameter to be shared across the entire frequency band. A partially known calibration standard is then a function of frequency (ω), as well as the unknown parameters (φ).

$$\mathbf{a}_i\left(\omega, \varphi\right) = f\left(\omega, \varphi_0, \varphi_1, \ldots \varphi_p\right) \tag{3.8}$$

Consequently, the residual error to be minimized is summed across all standards and frequencies simultaneously.

$$\epsilon^2 = \sum_{\omega}^{W} \sum_{i}^{I} \left| \mathbf{a}_i\left(\omega, \varphi\right) - f^{-1}\left(\mathbf{m}_i\left(\omega\right); E\left(\omega\right)\right) \right|^2 \tag{3.9}$$

Where I is the number of standards and W is the number of frequency points

3.4. Implementation

To illustrate the theory given above, an application to one-port calibration is given below. For computational reasons, the implementation contains a mixture of linear and non-linear methods. The parameters of the embedding circuitry are found through direct matrix inversion at each frequency (linear), while the unknown parameters of the calibration standards are found through an iterative method (nonlinear). The mixture of linear and non-linear methods is not essential to the concept, but was used for algorithmic simplicity.

3.4.1. Conventional One-port Calibration

The measured reflection coefficient of a one-port network embedded behind some unknown two-port is

$$\mathbf{m} = e_{00} + \frac{e_{01}e_{01} \cdot \mathbf{a}}{1 - e_{11} \cdot \mathbf{a}} \tag{3.10}$$

Where

$$\mathbf{m} = \text{measured reflection coefficient}$$

$$\mathbf{a} = \text{actual reflection coefficient}$$

$$\begin{bmatrix} e_{00} & e_{01} \\ e_{10} & e_{11} \end{bmatrix} = \text{embedding network s-parameters}$$

The conventional approach to one port calibration is to convert the nonlinear func-

tion (2.1) into a linear one. This is done by treating the determinant of the embedding network s-parameter matrix as a single variable ($\Delta e = e_{00}e_{11} - e_{01}e_{10}$) and mixing the independent and dependent variables to produce the following,

$$e_{11}\mathbf{m}\mathbf{a} - \Delta e\,\mathbf{a} + e_{00} = \mathbf{m} \tag{3.11}$$

Given three raw measurements ($\mathbf{m}_{1\ldots3}$) of calibration standards with known responses ($\mathbf{a}_{1\ldots3}$), the unknowns ($e_{00}, e_{11}, \Delta e$) can be directly found using matrix inversion.

$$\underbrace{\begin{bmatrix} \mathbf{m_1}\cdot\mathbf{a_1} & -\mathbf{a_1} & 1 \\ \mathbf{m_2}\cdot\mathbf{a_2} & -\mathbf{a_2} & 1 \\ \mathbf{m_3}\cdot\mathbf{a_3} & -\mathbf{a_3} & 1 \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} e_{11} \\ \Delta e \\ e_{00} \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{m_1} \\ \mathbf{m_2} \\ \mathbf{m_3} \end{bmatrix}}_{\mathbf{b}} \tag{3.12}$$

$$\begin{aligned} \mathbf{A}\cdot\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &= \mathbf{A^{-1}}\cdot\mathbf{b} \end{aligned} \tag{3.13}$$

To employ partially known calibration standards, additional measurement information is required, meaning more than three calibration standards must be measured. This creates an overdetermined system of equations, from which a residual error may be calculated. The unknown parameters of the partially known calibration standards are then found by minimizing this residual error.

Given additional measurements, (3.12) becomes overdetermined. The least squares solution to such a system is found by calculating the pseudo-inverse of $\mathbf{A}$, given by $\mathbf{A}^+ = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$ [17]. Where the asterisk ($*$) represents conjugate transpose. Thus, (3.12) becomes,

$$
\begin{aligned}
\mathbf{x} &= (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^* \cdot \mathbf{b} & (3.14)\\
&= \mathbf{A}^+ \cdot \mathbf{b} & (3.15)
\end{aligned}
$$

The residual error vector is given by ,

$$
\mathbf{e} = \mathbf{b} - \mathbf{A} \cdot \mathbf{A}^+ \cdot \mathbf{b} \tag{3.16}
$$

The residual error summed over all standards is calculated by,

$$
\epsilon = |\mathbf{e}|^2 \tag{3.17}
$$

Thus, the residual error of a overdetermined calibration provides the feedback mechanism needed to determine the unknown parameters of the calibration standards. Multi-frequency calibration is implemented by calculating the sum of the residual error magnitude across the entire frequency band. Because the unknown parameters are generally characteristics of the calibration standards, it is natural to first introduce the concept of parametric standards.

3.4.2. Parametric Standards

A parametric calibration standard has a response defined across a range of frequencies that can be modeled as a function of p adjustable parameters, $\varphi_{1...p}$.

$$\mathbf{a}_i\left(\omega, \varphi\right) = f\left(\omega, \varphi_0, \varphi_1, \ldots \varphi_p\right) \tag{3.18}$$

Thus, the system is constrained by assuming the calibration standard can be accurately represented by some electrical model. The model parameters ($\varphi_{1...p}$) generally correspond to some meaningful physical or electrical quantities that may be of interest. An example of a parametric standard would be a delayed termination of known reflection coefficient (Γ_0) and propagation constant ($\gamma\left(\omega\right)$) but unknown physical length, φ_1.

$$\mathbf{a}\left(\omega, \varphi\right) = \Gamma_0 e^{-2j\gamma(\omega)\varphi_1} \tag{3.19}$$

In this example, the response of the standard at all frequencies is determined by a single scalar parameter, φ_1. Replacing the fully known calibration standards in (3.12) with one or more parametric calibration standards will, in general, make (3.12) a non-linear system with respect to the parameters (φ). This system may be solved using various nonlinear least squares algorithms [47]. The quantity to be minimized is the squared magnitude of the residual error as given by (3.16), summed

across all calibration standards and frequencies.

$$min(\sum_{\omega} \left|\epsilon\left(\omega\right)\right|^2 ; \varphi) \qquad (3.20)$$

3.4.3. Convergence

It is important to recognize that the convergence of this technique is dependent upon the nature of the residual error function, which must be well-behaved and possess a unique minimum within a reasonable parameter space about the true parameter values. In general, it is not be possible to analytically prove the existence of such a minimum. A numerical simulation of the residual error function, demonstrated in Section sec. 3.5.2, is a convenient alternative when the parameter space is small. As the parameter space gets larger, Monte-Carlo simulations must be performed to determine the solvability a given parameter space[28, 19].

3.5. Application

3.5.1. Delay Shorts of Unknown Length

A common set of calibration standards used in waveguide measurements is a flush short, one or more delay shorts, and a load. At THz frequencies, these standards become increasingly difficult to accurately fabricate and characterize. This, along with waveguide misalignment, severely limits measurement accuracy in this medium[6]. Currently, there are efforts to evaluate competing design variations of traditional standards[10], introduce novel standards[12], and improve alignment precision by creation of new flange specifications[11].

While these efforts toward higher precision and lower tolerances on calibration standards are justified, imperfect hardware will always exist at the forefront of high frequency technologies. The solution is to replace the requirement for perfect hardware with the intelligent use of imperfect hardware. Multi-frequency self-calibration is a ideally suited for this task. For example, a common problem with delay short shims at WR-1.5(WM-380) is that their nominal delay lengths are in error. This may be due to incorrect characterization, machining tolerances, or possibly an imperfect test-port. To address this problem, a self-calibration utilizing parametric standards for delayed shorts with unknown lengths may be utilized.

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be parametric offset shorts of unknown lengths φ_1 and φ_2, but known propagation constant, $\gamma(\omega) = j\alpha(\omega) + \beta(\omega)$. Their complex reflection coefficients are thus,

$$\mathbf{a}_1(\omega, \varphi_1) = -e^{-2j\gamma(\omega)\varphi_1}$$

$$\mathbf{a}_2(\omega, \varphi_2) = -e^{-2j\gamma(\omega)\varphi_2}$$

Let $\mathbf{a}_3$ be a waveguide short,

$$\mathbf{a}_3(\omega) = -1$$

And, Let $\mathbf{a}_4$ be some known load standard.

$$\mathbf{a}_4(\omega) = \Gamma_l(\omega)$$

Following equations (3.12) - (3.20) for the above calibration set, the system of equations is

$$
\underbrace{\begin{bmatrix}
-\mathbf{m}_1\mathbf{a}_1(\omega,\varphi_1) & -\mathbf{a}_1(\omega,\varphi_1) & 1 \\
-\mathbf{m}_2\mathbf{a}_2(\omega,\varphi_2) & -\mathbf{a}_2(\omega,\varphi_2) & 1 \\
-\mathbf{m}_3 & 1 & 1 \\
\mathbf{m}_4{\cdot}\boldsymbol{\Gamma}_1 & -\boldsymbol{\Gamma}_1 & 1
\end{bmatrix}}_{\mathbf{A}}
\cdot
\underbrace{\begin{bmatrix}
e_{11} \\
\Delta e \\
e_{00}
\end{bmatrix}}_{\mathbf{x}}
=
\underbrace{\begin{bmatrix}
\mathbf{m}_1 \\
\mathbf{m}_2 \\
\mathbf{m}_3 \\
\mathbf{m}_4
\end{bmatrix}}_{\mathbf{b}}
\tag{3.21}
$$

The residual error is calculated by (3.16) and (3.17). The lengths of the delay lines are found by minimizing this residual error across all frequencies,

$$
min(\sum_{\omega}|\epsilon\left(\omega\right)|^2;\varphi)
\tag{3.22}
$$

Once the calibration standard's unknown parameters are found, the scattering parameters of the embedding circuit are found from (3.14).

3.5.2. Numerical Simulation of Residual Error Function

As stated in sec. 3.5.2, a numerical simulation should be constructed to verify the existence of a unique minimum in the residual error function within a reasonable parameter space. The calibration kit defined in the previous section was simulated in the 500-750GHz (WR-1.5, WM-380) band and contained the following standards:

- flush short,

- $\frac{\lambda}{4}$ delay short

- $\frac{\lambda}{8}$ delay short

- load (waveguide radiating open)

The true responses of the standards were embedded in a realistic two-port error network, producing simulated measurements. This realistic error network was taken from a previous conventional one-port calibration made in the laboratory. The actual lengths of the two offset shorts were then perturbed independently by known amounts from their nominal values. For each perturbation of the standards, the least squares estimate of embedding network's s-parameters was computed from (3.14) and the residual error was calculated according to (3.22). Because the parameter space is two dimensional it can be easily visualized as a surface. Although not a rigorous proof of solvability, the visualization of the residual error surface is conceptually useful. A surface with height proportional to the magnitude of the residual error function about a parameter space of $\pm 15\mu m, [\pm 7°]$ from the true values of φ_1 and φ_2, is shown below in Figure (Fig. 3.1). The boundaries of the parameter space ($\pm 15\mu m, [\pm 7°]$) are chosen to be larger than typical machining tolerances.

The fact that this surface is smooth and well behaved, indicates that convergence of the minimization is reasonable. Because no addition sources of error were included in this simulation, the error function exhibits a minimum value of 0 about the true values of the delay lengths, as expected.

3.5.3. Measurement Results

As a demonstration of the multi-frequency self-calibration routine described above, one-port measurements at 500-750GHz were made at the University of Virginia using a frequency extension unit from Virginia Diodes Inc. (VDI WR1.5 VNAXTXRX) with a Rhode and Schwarz ZVA 40 network analyzer as the backend. A calibration kit was obtained from VDI consisting of a flush short, four delay shims of two different nominal lengths, and a matched load. Initially, corrected measurements
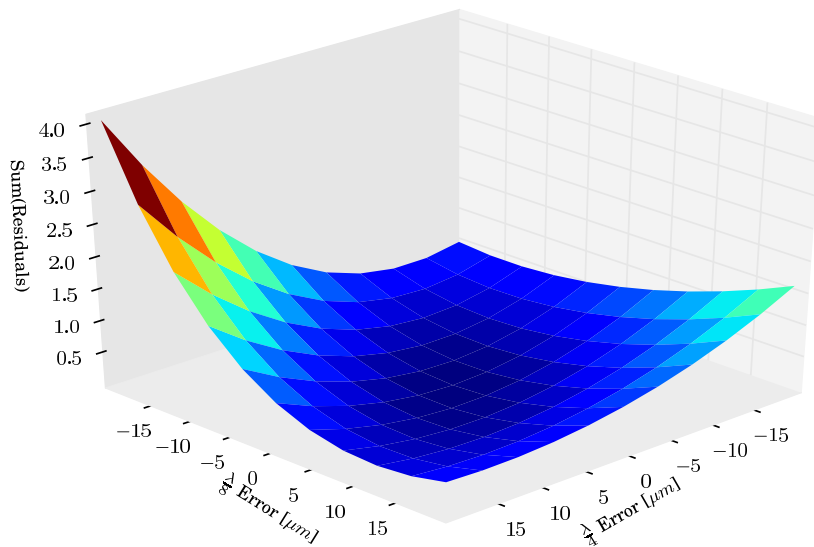
Figure 3.1.: Residual error function about true parameter values.

exhibited standing waves in their return loss, which is an indication that the load is not sufficiently matched to determine the system impedance. To circumvent this problem, the matched load was replaced by a radiating open waveguide flange. As explained in [7], this standard is also attractive because it is resistant to waveguide flange misalignment. The modeled response of the open waveguide was generated with HFSS simulations, and has since been studied in greater detail[12].

The results of the parametric self-calibration using delayed shorts of unknown length are compared to the conventional, least-squares calibration using the delay short's nominal lengths. Shown in Figure Fig. 3.2 is the evolution the self-calibration's residual error as a function of iteration index of the minimization routine. The minimization was accomplished using the `fmin` function of Python module `scipy.optimize`[48], which is an implementation of the Nelder-Mead simplex algorithm[49]. A comparison of the corrected delay short measurements calibrated using the nominal lengths vs those calibrated with the corrected lengths found by optimization is shown in Figure Fig. 3.3. The results suggest that the true delay lengths are significantly dif-

ferent from their nominal values, and the self-calibration routine provides improved corrected measurement accuracy.
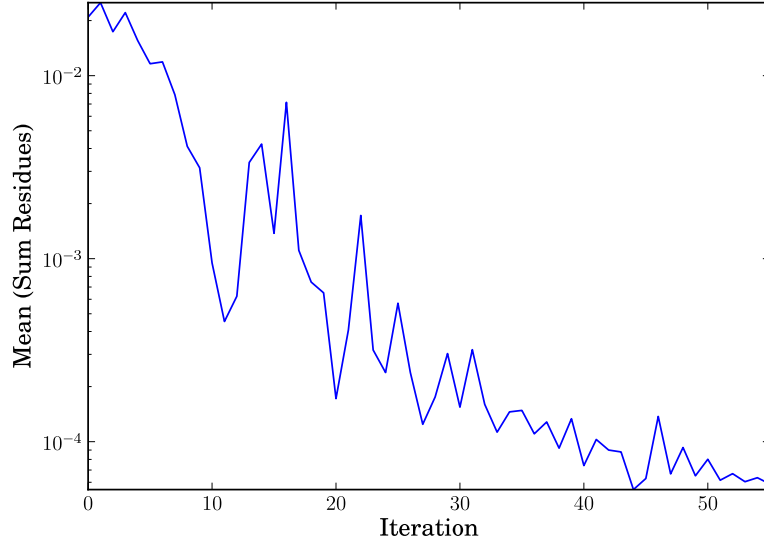
Figure 3.2.: Residual error vs iteration of minimization algorithm.

Corrected measurements of the load standard were then calculated with the optimal calibration set. It was observed that the load exhibited a significant reflection electrically far from the flange. It is known that load standard is constructed by inserting a piece of absorber in a section of waveguide, and therefore, this result is consistent with the physical model. It was hypothesized that the match could be modeled as a flat-band reflection, located at some fixed distance from the reference plane.

$$\mathbf{a_i}\left(\omega, \varphi_0, \varphi_1\right) = \varphi_0 \cdot e^{-2j\gamma(\omega)\cdot\varphi_1}$$

Where φ_0 is the magnitude of the reflection coefficient, and φ_1 is the physical distance to the absorber. Thus, the fully characterized match standard was replaced
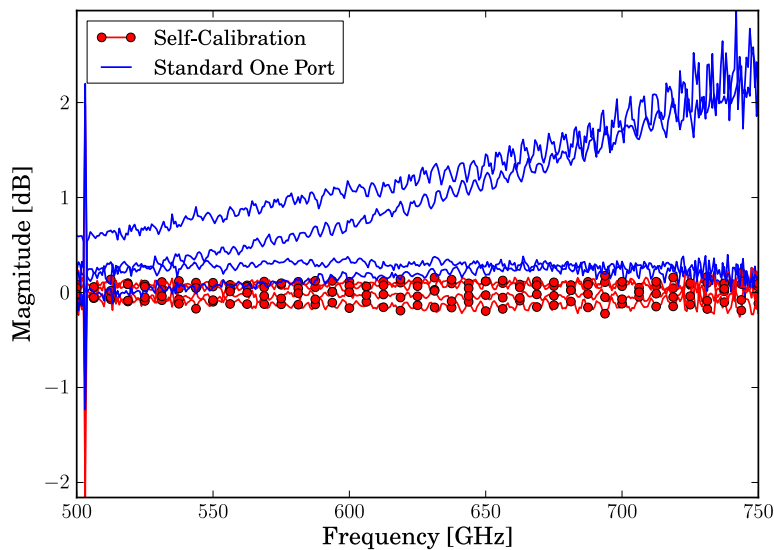
Figure 3.3.: Comparison of delay-shorts reflection coefficient magnitude, as corrected with conventional and self-calibration algorithms.

by a partially known standard, parameterized by delay length and reflection coefficient magnitude. Using the previous calibration, initial parameter estimates for the delayed-load were determined. The results of the self-calibration which used the parametric load standard were then compared to a calibration that used the optimal delay-short lengths, but an ideal match for a load. As a performance metric, the corrected measurements of the short and two delay shorts are shown in Fig. 3.4, and similarly for the load in Fig. 3.5.

The return loss of the delayed shorts calibrated with the refined self-calibration are below $\pm.1$dB deviation for most of the band, while the initial self-calibration produces results around $\pm.5$dB. The self-calibration finds that the absorbing load has a -37dB return loss, at 1.9cm (.75 in) from the reference plane, so Figure Fig. 3.5 should be interpreted accordingly. An iterative approach to self-calibration is useful in that it provides a practical way to deal with poorly characterized standards, a current problem for accurate millimeter-wave and sub-millimeter-wave measure-
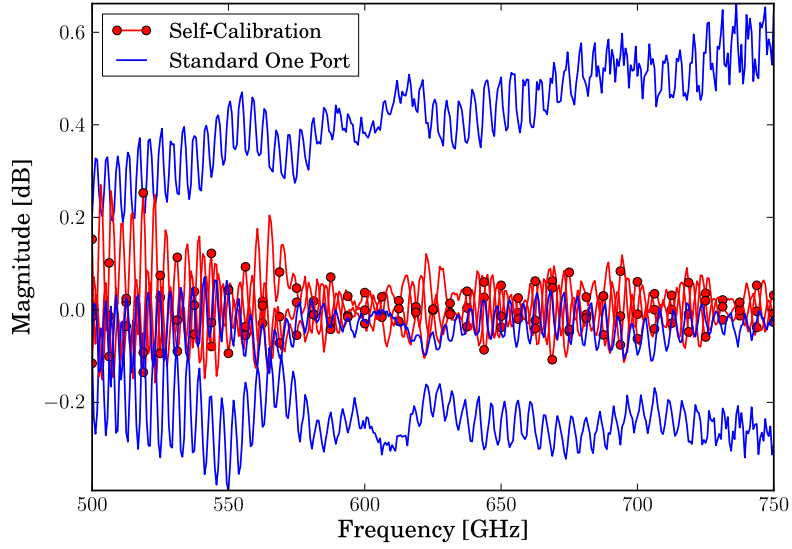
Figure 3.4.: Comparison of delay-shorts reflection coefficient magnitude, as corrected with initial and refined self-calibration algorithms.

ments.

3.6. Conclusions

A general approach to multi-frequency, one-port self-calibration using parametric standards has been presented. A specific algorithm, utilizing an iterative least squares approach is described and demonstrated through practical examples of waveguide calibration using delay shorts and delayed loads of unknown lengths. The self-calibration technique shows improvement over conventional one-port calibration, given the same initial standard definitions. This specific application is of importance to the emerging THz measurement infrastructure, where rectangular waveguide calibration limits measurement accuracy. More sophisticated parametric standards could be developed to further improve accuracy. For example, to circumvent the waveguide misalignment at the test-port caused by machining tolerances,
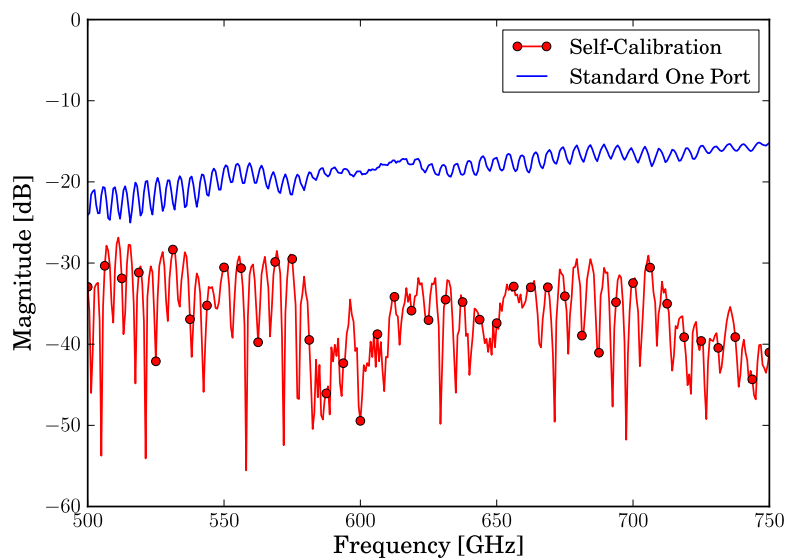
Figure 3.5.: Comparison of the matched load standard's reflection coefficient magnitude, as corrected with initial and refined self-calibration algorithms.

one could account for the electrical effects of the misalignment by parameterizing the offset of the waveguide junction. Extending this technique to two-port calibration, or 2n-port calibration, could be accomplished in a variety of ways. The implementation of the formulation described in [22], which relies on the 'matricial bilinear transform', is underway.

4. An Experimental Technique for Calibration Uncertainty Analysis

This chapter presents an experimentally-based technique for characterizing calibration uncertainty. The approach described calculates uncertainty metrics at the output of the calibration processing chain as opposed to the input. In doing so, this method replaces the complexities of error propagation with the computational effort associated with performing numerous calibrations. Practical applications are demonstrated for a variety of scenarios focused on one-port calibration, illustrating the versatility of the technique. An Open-Source implementation has been made publicly available as part of the Python module scikit-rf.

4.1. Introduction

Uncertainty estimation of vector measurements at microwave frequencies is an important and well developed subject [50, 27, 51, 19]. Statistical methods are regularly used to propagate uncertainties in calibration standards to uncertainties in measured quantities. Software tools exist, and more are under development [52]

that can be used to aid the experimentalist in uncertainty estimation. These techniques and tools, although powerful, are sometimes prohibitively complex for the average experimentalist. A simple, transparent, and easy-to-implement procedure for estimating uncertainty in S-parameter measurements is desirable.

The typical approach to estimating calibration uncertainty is to calculate uncertainty metrics on measurements of calibration standards, and then to propagate those metrics through the calibration processing chain. Modern approaches accomplish this through the use of variance-covariance and Jacobian matrices [27, 29]. There are several drawbacks of this approach to uncertainty estimation, as are described in [50].

First, in the context of calibration, error propagation relies on a linear approximation to a non-linear function. The magnitude of the non-linearity is a function of the embedding network, and may be quite large. Second, implementation requires the evaluation of partial derivatives, which is a non-trivial task. Finally, error propagation is incapable of accurately representing certain types of unbiased errors (see Section sec. 4.4.3 for an example).

The approach described here is different in that the uncertainty metrics are calculated at the output of the calibration processing chain instead of the input. Thus, the complexity of error propagation is replaced by the computational effort required to perform numerous calibrations. In a sense, this is a step backward in mathematical sophistication, but provides a useful procedure that produces intuitive results.

The replacement of error propagation with Monte-Carlo analysis was presented in [50] as a general way to estimate uncertainty in quantities related to S-parameters. The method presented here is essentially the application of Monte-Carlo analysis to experimentally measured data. Measurement-based methods have the advantages that no assumptions are made about the error distribution and all unbiased

uncertainties are automatically taken into account, as opposed to only predicted ones.

It should be observed that this type of uncertainty estimation quantifies the effects of unbiased errors only, and not biased errors. In other words, it is an estimation of measurement precision, not measurement accuracy.

4.2. Theory

This section follows the introduction to error propagation as presented in [53], and draws from the language and concepts as described in [54]. Assume there is some measurand x, that is calculated from measurements $(u, v, ...)$, by way of a measurement equation, f.

$$x = f(u, v, \ldots) \tag{4.1}$$

Provided that redundant measurements are made, multiple values for $u_i, v_i, ...$ are available. By combining individual measurements of $u_i, v_i, ...$, a set of measurands $\{x_i\}$, can be generated.

$$\{x_i\} = f\Big(\{u_i\}, \{v_i\}, ...\Big) \tag{4.2}$$

The uncertainty in the value of x can then be calculated directly from the set $\{x_i\}$. This eliminates the need for error propagation, and replaces it with redundant evaluations of the function f.

Applying this concept to the problem of calibration is straightforward, but will be explicitly expressed for clarity. Given redundant measurements of the calibration standards, numerous calibrations may be performed that generate a set of error networks $\{E_i\}$, or similar set of error correction information[1]. This set of error networks is a function of the *assumed* response for calibration standards a, and the set of redundant measurements of the calibration standards $\{s_i\}$.

$$\{E_i\} \;=\; f\left(a, \{s_i\}\right) \tag{4.3}$$

A set of corrected responses, $\{c_i\}$, can be produced by de-embedding the set of error networks from a single measurement, m. Thus, the set of corrected responses is a function of a single measured response and a set of error networks.

$$\{c_i\} \;=\; f\left(m, \{E_i\}\right) \tag{4.4}$$
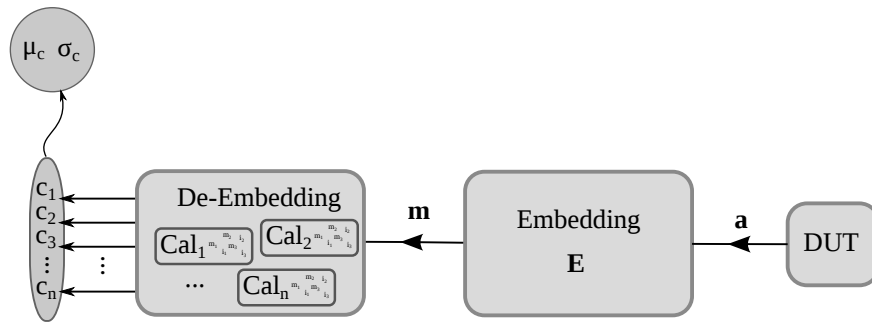
Figure 4.1.: Illustration of of calibration uncertainty algorithm

---

[1]The form of the error-terms will vary depending on what error model is used.

An illustration of the algorithm is shown in Fig. 4.1. The uncertainty of the corrected responses, or the error networks themselves, can be found directly from the sets $\{c_i\}$ and $\{E_i\}$, respectively. If the uncertainty in some other related quantity is desired, then the functional relationship can be incorporated into the processing chain in a similar manner. Although previously stated, it should be emphasized that uncertainties estimated in this way will capture the effects of unbiased errors only.

The subscripts on set variables have been kept the same for simplicity of presentation, ie $f\Big(\{u_i\}, \{v_i\}, ...\Big)$, but it is not necessary that the indicies of the measurand and input variables be identical. In fact, the freedom to choose which combinatoric function is used to assemble the elements of each set contributes to the versatility of the method (this is demonstrated in Section sec. 4.4.3).

To illustrate the theory, an application to a conventional one-port calibration algorithm is given in the following section.

4.3. Implementation

The measured reflection coefficient of a one-port device embedded in a two-port error network is given by the familiar formula,

$$\mathbf{s} = e_{00} + \frac{e_{01}e_{01} \cdot \mathbf{a}}{1 - e_{11} \cdot \mathbf{a}} \tag{4.5}$$

where

$\mathbf{s} =$ measured reflection coefficient

$\mathbf{a} =$ actual reflection coefficient

$$\begin{bmatrix} e_{00} & e_{01} \\ e_{10} & e_{11} \end{bmatrix} =$$ embeding network S-parameters

A common approach to one port calibration is to convert the nonlinear function (2.1) into a linear least squares problem. This is done by treating the determinant of the embedding network as a single variable, Δe, and mixing the independent and dependent variables to produce the following,

$$e_{11}\mathbf{sa} - \Delta e\,\mathbf{a} + e_{00} = \mathbf{s} \tag{4.6}$$

If one measures three calibration standards, each having a known response, the unknowns $(e_{11}, \Delta e, e_{00})$ can be uniquely found through matrix inversion.

$$\underbrace{\begin{bmatrix} \mathbf{s_1 a_1} & \mathbf{a_1} & 1 \\ \mathbf{s_2 a_2} & \mathbf{a_2} & 1 \\ \mathbf{s_3 a_3} & \mathbf{a_3} & 1 \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} e_{11} \\ -\Delta e \\ e_{00} \end{bmatrix}}_{\mathbf{E}} = \underbrace{\begin{bmatrix} \mathbf{s_1} \\ \mathbf{s_2} \\ \mathbf{s_3} \end{bmatrix}}_{\mathbf{S}} \tag{4.7}$$

$$\begin{aligned} \mathbf{A} \cdot \mathbf{E} &= \mathbf{S} \\ \mathbf{E} &= \mathbf{A^{-1}} \cdot \mathbf{S} \end{aligned} \tag{4.8}$$

76

This inversion yields an estimate for the embedding network s-parameters, allowing one to remove the effects of the embedding network on future measurements. The corrected response $\mathbf{c}$ of a future measurement $\mathbf{m}$ is calculated by inverting (2.1), and inserting values of $(e_{11}, \Delta e, e_{00})$ as found by (4.8)

$$\mathbf{c} = \frac{\mathbf{m} - e_{00}}{e_{11}\mathbf{m} + \Delta e} \tag{4.9}$$

Thus, the corrected response $\mathbf{c}$ is a function of it's measurement $\mathbf{m}$, as well as the estimated parameters of the embedding network $\mathbf{E}$. The estimated parameters of the embedding network are dependent upon the calibration standards' assumed and measured responses through (4.8). This relationship may be expressed by,

$$\mathbf{c} = f\left(\mathbf{m}; \mathbf{E}\right) = f\left(\mathbf{m}; \mathbf{a}, \{\mathbf{s}\}\right) \tag{4.10}$$

The uncertainty in the corrected response due to variance in the calibration standard measurements can be estimated by evaluating equations (4.8) and (4.9) repeatedly with redundant measurements of the calibration standards $\{\mathbf{s}\}$. This produces a set of corrected responses $\{\mathbf{c}\}$, on which uncertainty bounds may be calculated.

Although not demonstrated in this study, the effects of other uncertainties can be calculated just as easily. For example, the uncertianty in the calibrated response due to the variance in the assumed responses of the calibration standards could be estimated. This may be used to estimate the effects of calibration standard tolerances on corrected measurements. The technique described above has been implemented as part of the Object-Oriented, Open-Source Python Module scikit-

rf[38]. The measured and ideal network data is taken in the form of Touchstone files, the standard file format for representing microwave networks and supported by all modern Vector Network Analyzers[55]. For more details on using scikit-rf for uncertainty estimation, see the on-line documentation [38].

4.4. Applications

This section presents examples of calculating measurement uncertainties in three different calibration media, each with a different source of uncertainty.

4.4.1. Rectangular Waveguide

The uncertainty analysis method described has been used to assess different waveguide calibration standard designs for the 500-750GHz band (WM-380 or WR-1.5). Measurements were made using a one-port WR-1.5 frequency extension unit from Virginia Diodes Inc. (VDI WR1.5 VNAXTXRX) with a Rhode and Schwarz ZVA-40 network analyzer as the back-end. A full account of the dominant error sources and measurement challenges of this medium are not trivial and are the subject of on-going research[6]. The largest limitation of measurement accuracy at this frequency band is flange misalignment of the calibration standards.

Waveguide delay shorts are common calibration standards in this medium, and can be realized as a recess milled into a flanged short or as a thin shim backed by a flush short. The two different designs are illustrated in Figure Fig. 4.2. Shim-based standards are attractive because they have dual use as delay lines and delay shorts, but they have larger misalignment uncertainty than the milled design[56]. To quantify the difference between the two standards, the calibration uncertainty was estimated for calibrations containing each.
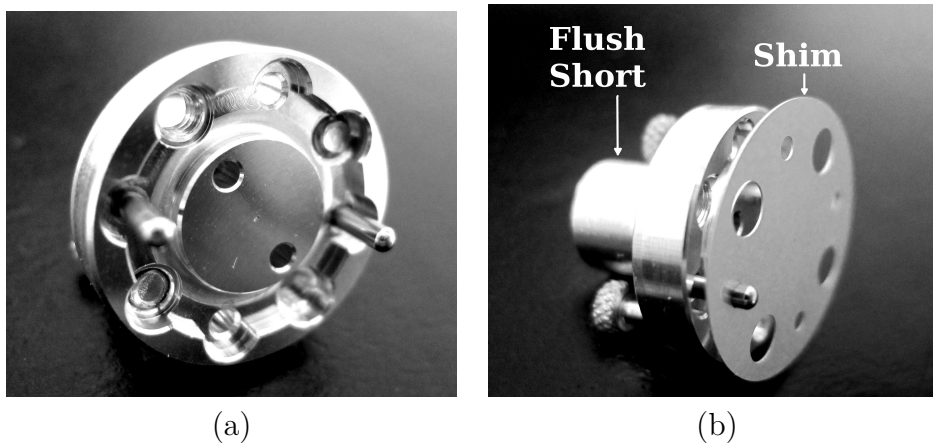
Figure 4.2.: Photographs of the two delay short designs; a) Milled design, and b) shim design (with accompanying flush short)

Two calibration sets were made, both containing a short, radiating open, and a quarter-wave delay short. One calibration set used the milled delay short, while the other used the shimmed delay short. The use of the radiating open as a load standard is discussed in [6, 7]. Six measurements were made of each standard. Between each individual measurement, the standard was removed and replaced on the test port. Thus, the variance of each standard reflects the uncertainty associated with the flange connection.

The measurement of a verification standard, an eighth-wave delay short, was corrected using both calibration sets. A direct comparison of the uncertainty produced by each calibration set is made by calculating the complex standard deviationof an eighth-wave delay short, as corrected by each set. The complex standard deviation of a set of points is the average distance of each point from the set's mean, expressed by,

$$\sigma\{x\} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|x_i - \frac{1}{n}\sum_{i=1}^{n}(x_i)\right|^2} \qquad (4.11)$$

This is shown in Figure Fig. 4.3. The standard deviation of the shim set varies from .03 to .14 across the band, while the milled set is below .02 for the entire band, with a mean value of .004.
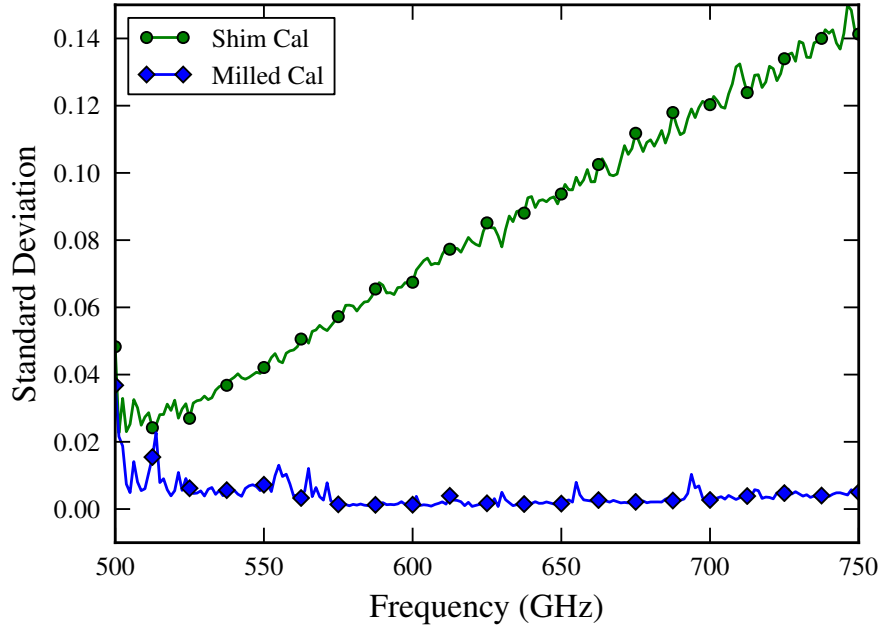
Figure 4.3.: Complex Standard Deviation of Eighth-wave Verification Standard For Calibration Sets Using Shimmed Delays vs. Milled Delays

Although the metric illustrated in Figure Fig. 4.3 provides a direct comparison of the uncertainty produced by each calibration set, a plot of the magnitude and phase components of the verification standard with confidence bounds is desirable. Because the entire set of corrected measurements is available, statistics on the magnitude and phase components can be calculated directly.

When calculating the uncertainty in the magnitude component to be displayed on a log scale, the bounds must be calculated in linear units, and then converted to log units. The order is important. For example, to calculate the $\pm 3\sigma$ uncertainty bounds in the magnitude in dB from a complex set $\{x\}$,

$$
\begin{aligned}
|x|_{upper}\,(\text{dB}) &= 20\log\left(\mu\left\{|x|\right\} + 3\sigma\left\{|x|\right\}\right)\\
|x|_{lower}\,(\text{dB}) &= 20\log\left(\mu\left\{|x|\right\} - 3\sigma\left\{|x|\right\}\right)
\end{aligned}
\tag{4.12}
$$

Where μ is the mean of the set, and the absolute value operates on each element of the set. The magnitude and phase components, along with $\pm 3\sigma$ confidence bounds for each calibration set, are shown in Figures Fig. 4.4 and Fig. 4.5. The mean response and confidence bounds for the shim delay short calibration set are indicated by square markers and light shading, while the mean response and confidence bounds for the milled set are indicated by diamond markers and dark shading, respectively. Although the mean responses are very similar, measurement uncertainty for the milled set is significantly smaller than the shim set. The insets of Figures Fig. 4.4 and Fig. 4.5 better illustrate the confidence bounds produced by the milled calibration sets. The $\pm 3\sigma$ confidence region for the milled set is less than .4dB across the band, while the $\pm 3\sigma$ confidence region for the shim set varies from $\pm.5$dB to ± 3dB. The difference between theory and measurement for the return loss phase as shown in Figure Fig. 4.5, is attributed to flange misalignment at the test port due to mechanical tolerances.

As pointed out in [6], the quadratic nature of the junction admittance created from flange misalignment in single direction, may produce biased errors in the calibration. Errors of this nature are not accounted for with this type of uncertainty estimation.
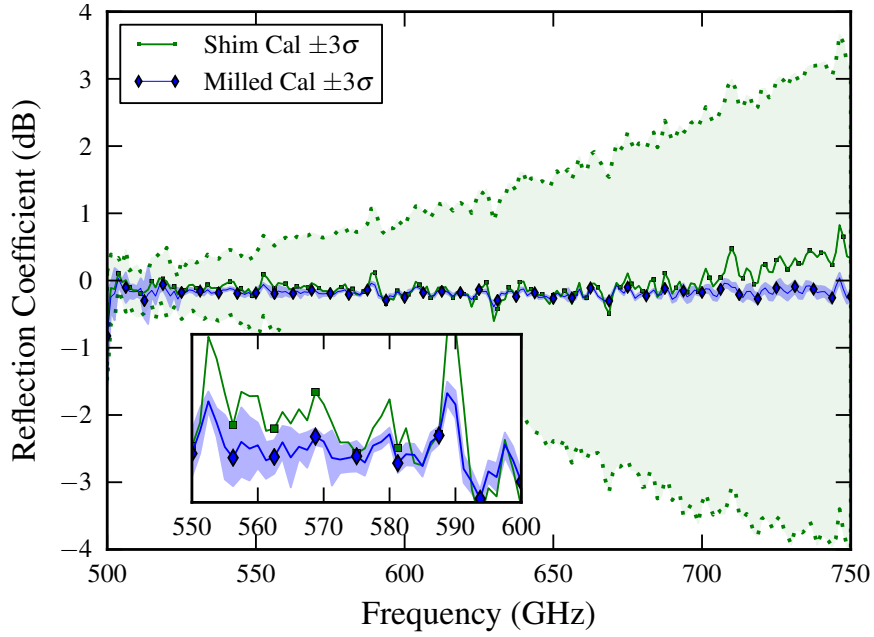
Figure 4.4.: Reflection Coefficient Magnitude of the Eighth-wave verification Standard with $\pm 3\sigma$ Confidence Bounds, Calibrated Using Shimmed Delays vs. Milled Delays. (The inset shows an enlargement to better illustrate confidence region of the Milled Set)

4.4.2. Two-Tier Coplanar Waveguide On-wafer Probe

As a second application, the S-parameters of a WR-1.5/WM-380 rectangular waveguide to coplanar waveguide probe were measured and evaluated using the experimentally based uncertainty analysis[5]. Because the probe is a reciprocal, two-port network it can be measured using a one-port, two-tier calibration, as described in [5]. The reference plane of the first calibration tier is the waveguide test port of the VNA. The second tier's reference plane is on the CPW side of the probe ($100\mu m$ past the probe's contact point on-wafer). A diagram illustrating the calibration tiers and positions with respect to the probe is shown in Figure Fig. A.10. A realistic rendering of the waveguide test-port, probe, and calibration substrate is shown in Figure Fig. 4.7.

The first-tier calibration was made at the waveguide test-port and employed a flush
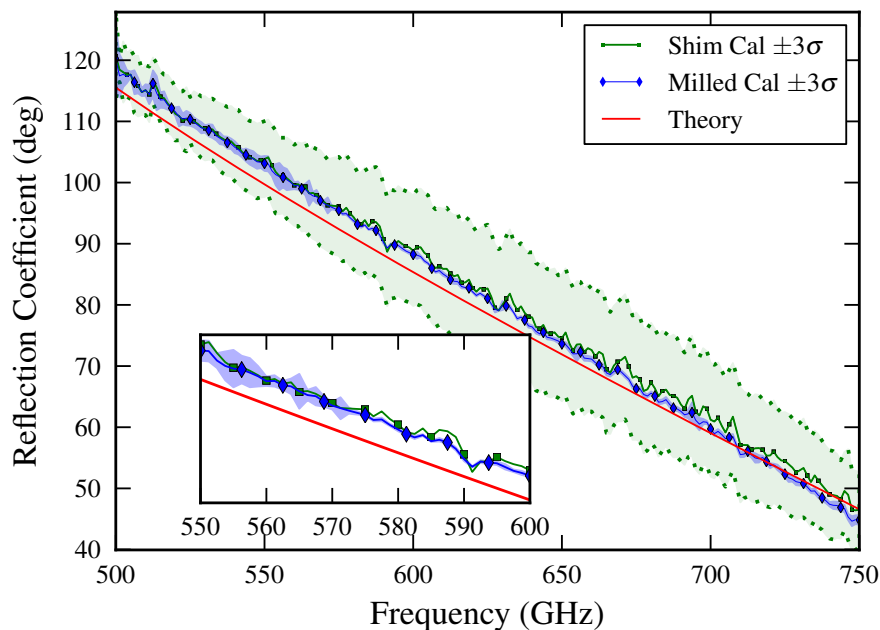
Figure 4.5.: Reflection Coefficient Phase of the Eighth-wave verification Standard with $\pm 3\sigma$ Confidence Bounds, Calibrated Using Shimmed Delays vs. Milled Delays. (The inset shows an enlargement to better illustrate confidence region of the Milled Set)

short, quarter-wave delay short, and radiating open as calibration standards. The second-tier calibration was made on the substrate and contained a CPW short and a set of four delayed shorts. The electrical lengths of the delays were evenly distributed from $0 - 180°$ at band-center, excluding the end-points. An illustration of the CPW calibration set and it's relevant geometry is shown in Figure Fig. 4.8. Using the error networks produced by both tiers, the S-parameters of the probe can be determined. In this scenario, the quantity of interest is the error network of the second tier calibration, after the VNA has been de-embedded from port 1.

An uncertainty analysis was made on the second tier calibration only. Each standard was measured 30 times. These measurements were combined to produce a set of 30 estimates for the S-parameters of the probe. Uncertainty estimates of the probe's insertion loss were made in two different ways. One uncertainty estimate
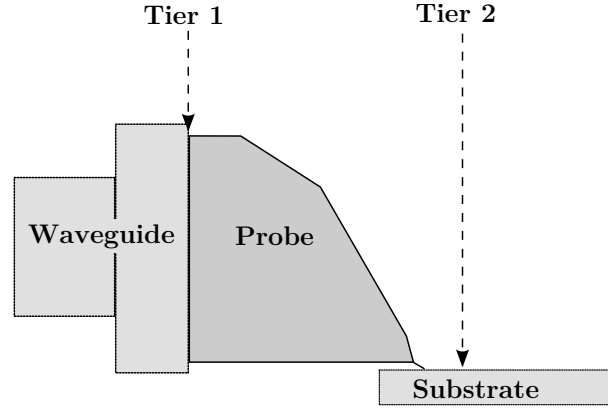
Figure 4.6.: Image of the Experimental Setup With the Calibration Tiers Labeled
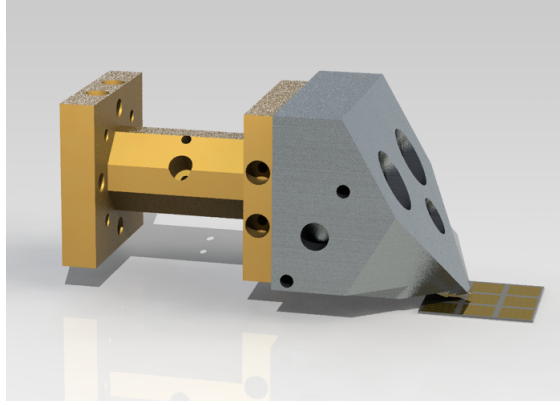
Figure 4.7.: Rendering of the Waveguide Test-port, Probe, and Calibration Substrate

was made by taking the standard deviation on the complex insertion loss $|\sigma\left(S_{21}\right)|$. The other estimate was made by taking the standard deviating of the magnitude of the insertion loss $\sigma\left(|S_{21}|\right)$. The difference between the two uncertainty estimators is due to the non-commutativity of statistical and projection operators, expressed by,

$$|\sigma\left(S_{21}\right)| \neq \sigma\left(|S_{21}|\right) \tag{4.13}$$

Both methods have unique interpretations; one is the uncertainty in the magnitude $\sigma\left(|S_{21}|\right)$ and the other is the magnitude of the complex uncertainty $|\sigma\left(S_{21}\right)|$.
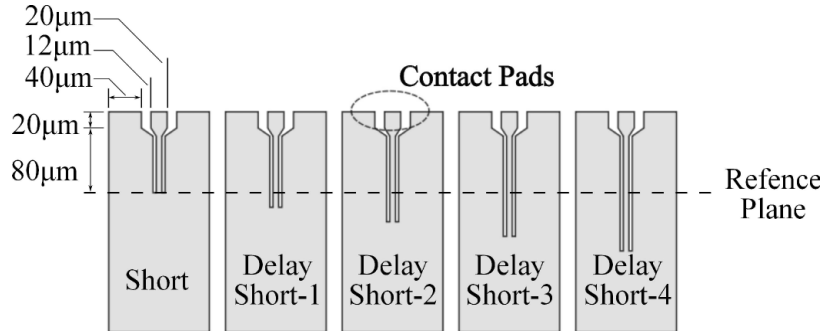
Figure 4.8.: Geometry of the second-tier CPW calibration lines Labeled

The mean insertion loss of the probe with $\pm 3\sigma$ uncertainty bounds are shown in Figure Fig. 4.9 for both uncertainty estimators. The mean responses are nearly identical, and their difference cannot be perceived from Figure Fig. 4.9. The bounds for the magnitude of the complex uncertainty $|\sigma(S_{21})|$, is shown by dotted lines and light shading, while the uncertainty in the magnitude $\sigma(|S_{21}|)$ is shown by dark shading. The difference in the two uncertainty estimators is significant, and interpretation is clear; the uncertainty in the insertion loss magnitude is very low, while the uncertainty in the magnitude of the complex uncertainty is much larger. The dominance of phase uncertainty in the transmission coefficient is attributed to the irregularity in probe landing possition. The 'spikes' seen in Figure Fig. 4.9 are numerical artifacts due to evaluation of insertion loss by way of the reciprocity assumption, which requires the square root of a complex number to be taken. These artifacts are only present for the uncertainty estimate that takes the magnitude component after evaluation of the standard deviation, i.e. $|\sigma(S_{21})|$.

4.4.3. Quasi-Optical Array

As a final application of this approach, the reflection coefficient of a quasi-optical annular ring-slot antenna array was measured in the 75-110GHz band (WM-2540
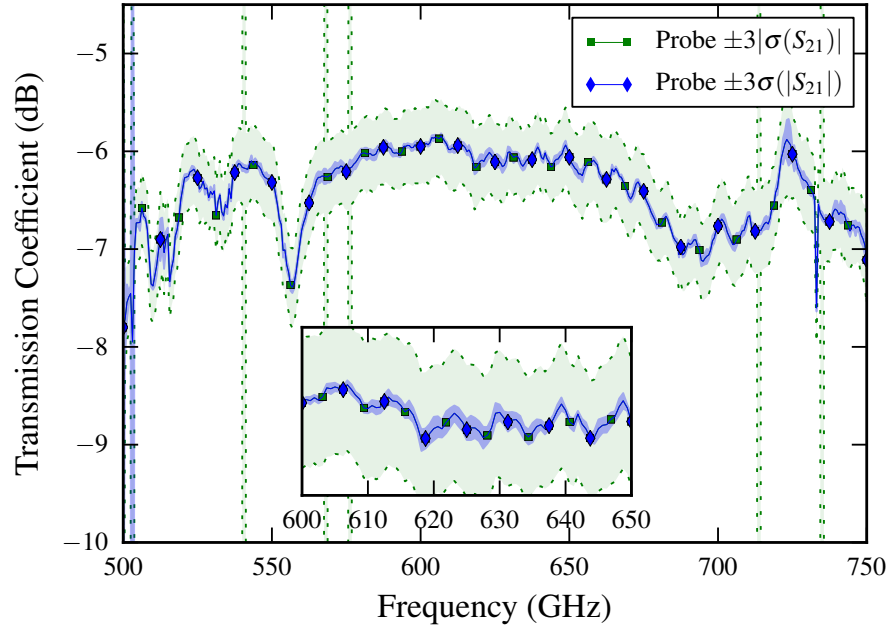
Figure 4.9.: Insertion Loss of Error Network With Uncertainty Estimation

or WR-10). The array was made up of a 16x16 square grid of 1.1x1.1mm unit cells. Each cell contained a ring slot of mean radius $400\mu m$ and slot width of $40\mu m$, fabricated on $100\mu m$ thick fused quartz wafer substrate ($\epsilon_r = 3.8$). An image of the array is shown in Figure Fig. 4.10.

A one-port calibration was performed using a series of delayed reflect standards, realized with a mirror and a mechanically adjustable optical path length. An illustration of the experimental setup is shown in Figure Fig. 4.11. The delay distance was controlled by an electronic translation stage, so that positioning accuracy and repeatability was excellent ($\sim 1\mu m$). The replacement of a device in the fixture, however, showed poor repeatability. Therefore, when the 'short' standard is replaced by the DUT, an uncertainty in location of the DUT with respect to the reference plane is introduced.

An analogous scenario is a coaxial or waveguide calibration that employs a sliding short only. In this case, the uncertainty due to the connector repeatability is present
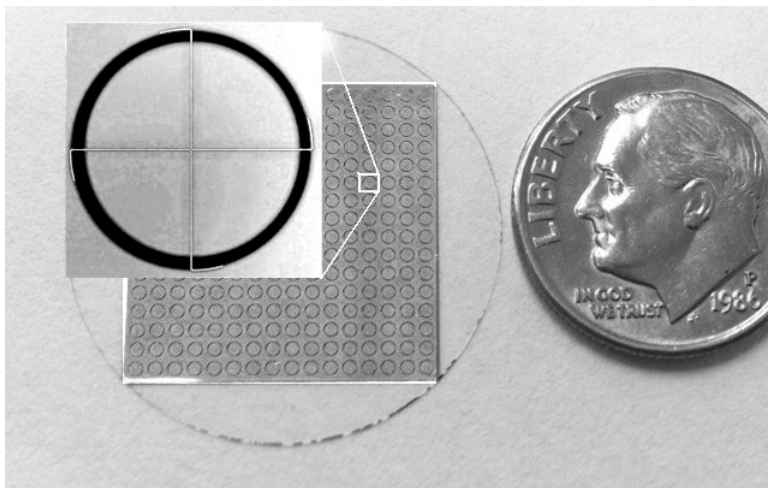
Figure 4.10.: Photograph of the Annular Ring-slot Array

when the calibration standard is replaced by the DUT, but not between the measurement of each calibration standard. Intuitively, one understands that the effect of this will be an uncertainty in the phase response of the DUT. However, accurately representing this type of uncertainty in the calibration is surprisingly subtle. For example, one way to estimate measurement uncertainty is to remove and replace the mirror from the fixture in between each measurement of a calibration standard. This will introduce a relative phase error in between the calibration standards themselves, and thus cause more significant calibration errors. Therefore, it may be said that the uncertainty estimate produced by this method would be pessimistic.

An alternative approach to calculate the uncertainty is to replace the mirror in the fixture only after each repetition of an entire set of calibration standard measurements, in which the mirror is translated but not re-mounted in the fixture. Representing this type of uncertainty is not possible with matrix-based error propagation techniques because such techniques require that uncertainty is formulated in terms of *per-standard* uncertainty.

The two uncertainty estimates described above were compared. One estimate was made from measurements that had the mirror replaced after the measurement of
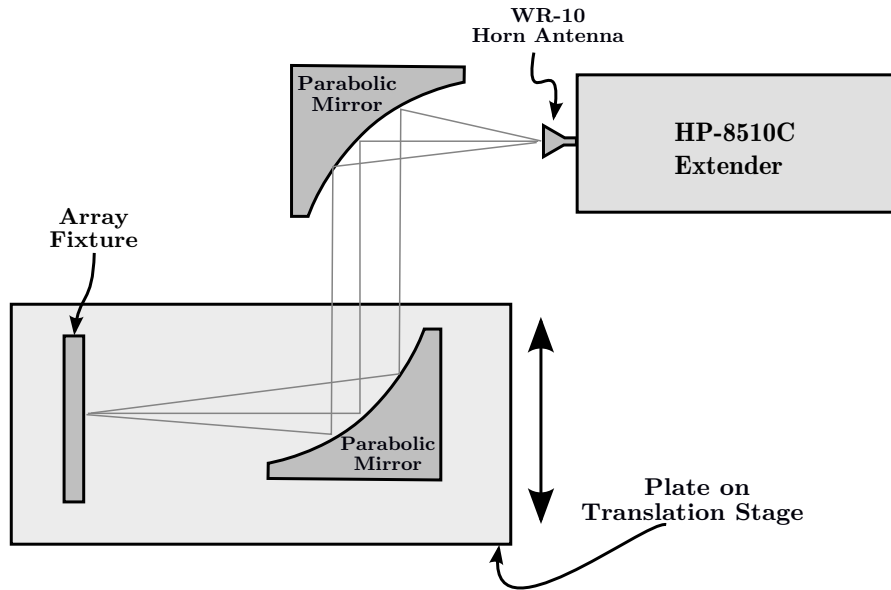
Figure 4.11.: Illustration of Quasi-optical Experimental Setup.

each standard, referred to as a *permuted* method, or *perm* for short. The other estimate has the mirror replaced after each set of standards is measured, referred to as the *dot* method. The names given to the uncertainty estimates are derived from the Python functions used to generate each set of corrected measurements. Both uncertainty estimates, as well as a simulated response of the array, are plotted in Figures Fig. 4.12-Fig. 4.14, for various components. The simulated response was created using Ansoft's High-Frequency Structure Simulator (HFSS).

As predicted, the *permuted* uncertainty estimate is more pessimistic than the *dot* estimate. Although the true response is never available to the experimentalist, the confidence bounds produced by the *dot* estimator very closely track the deviation of the measurement from simulation, allowing more confidence to be placed in the simulation model. The uncertainty in the phase component is not shown due to the numerical difficulties associated with calculating the variance of the phase component. Near resonance the array's phase component exhibits sharp jumps. These

sudden jumps produce difficulties in the unwrapping of the phase, a necessary step to express phase response uncertainty.
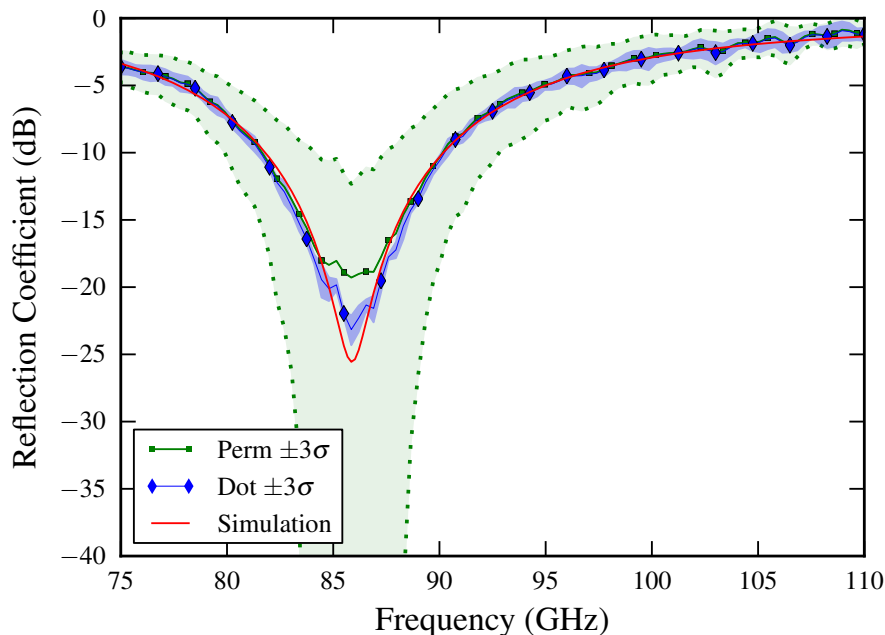
Figure 4.12.: Magnitude of Array Response Comparing Two Types of Uncertainty Estimates

4.5. Conclusion

We have presented an experimentally-based method of uncertainty estimation in which the uncertainty metrics are calculated at the output of the calibration processing chain as opposed to the input. This creates a conceptually simple and versatile procedure with results that are easy to interpret. The method has been implemented in the Object-Oriented, Open-Source Python module scikit-rf[38], and is made available on the Internet.

Examples of practical usage have been given for a variety of scenarios associated with one-port calibration. The extension of this technique to two-port calibration or any other measurement model is straightforward.
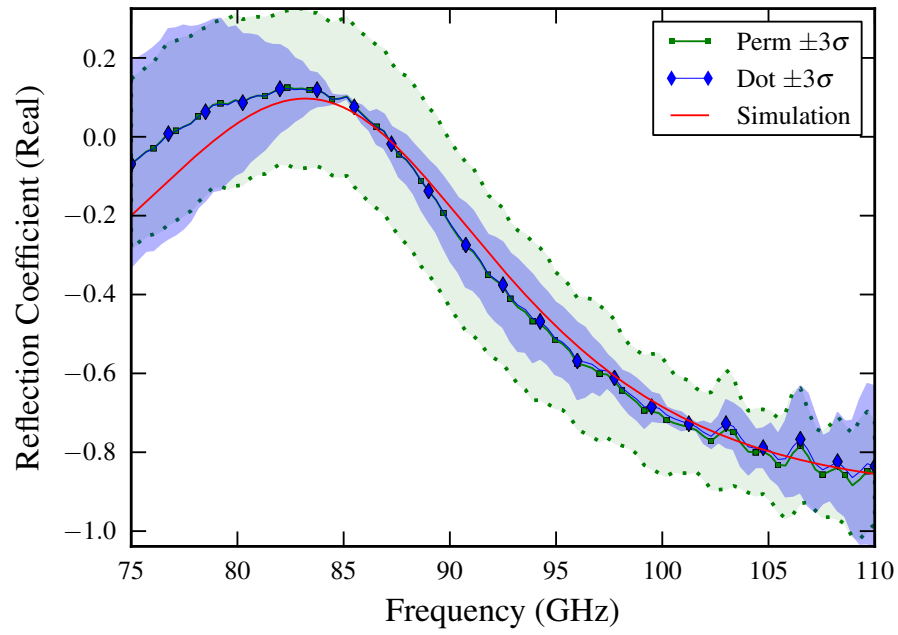
Figure 4.13.: Real Part of Array Response Comparing Two Types of Uncertainty Estimates
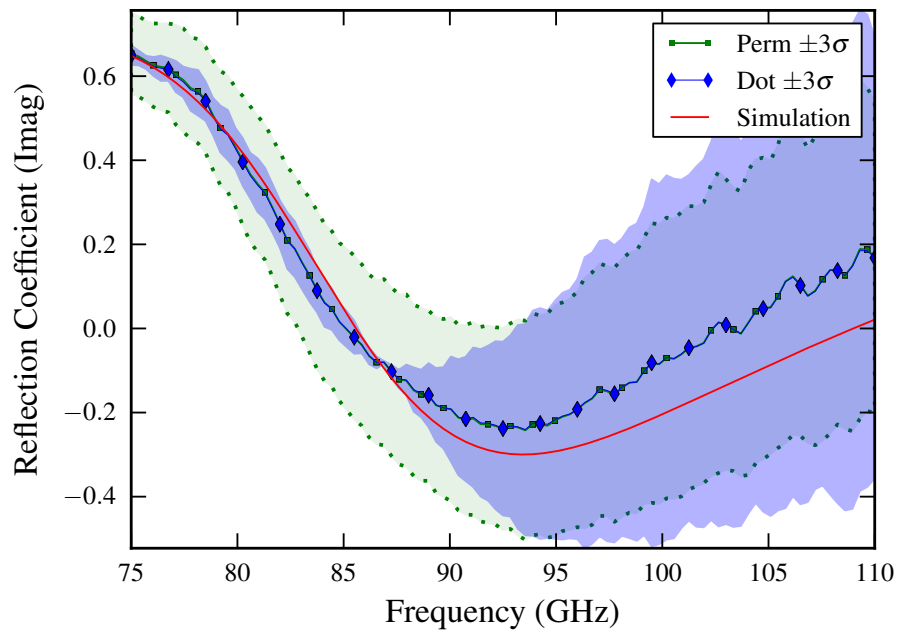
Figure 4.14.: Imaginary Part of Array Response Comparing Two Types of Uncertainty Estimates

5. `scikit-rf`: an Open Source Object Oriented approach to RF/Microwave Engineering

This chapter outlines some shortcomings in the current software available to the microwave research scientist. This provides the motivation for `scikit-rf`, an Open-Source, Object-Oriented Python module developed, in part, to support for the work in this thesis. Example scripts are provided that demonstrate how `scikit-rf` can be used to implement the applications described in chapter 3 and chapter 4.

5.1. The Problem

As computer software continues to play an increasing role in the field of engineering design and analysis, it's architecture becomes more critical. To a degree, it is the software's design that determines which solutions are chosen. The more refined and specialized the tool, the more pronounced it's effect becomes. This can create a problem for research-type environments, which constantly require new and

innovative solutions. This chapter reflects upon existing software tools from the perspective of a researcher, and attempts to identify their shortcomings so that they may be rectified.

Currently, numerous professional grade software packages exist to aid the microwave engineer in specific tasks. Shown in Tab. 5.1 is a partial list of some of the more popular software packages, sorted categorically. For those who are interested, a historical account of the evolution of circuit simulators written by Steve Maas can be found in [57].

EM Simulation	Circuitl Design	Calibration	Analysis
HFSS	ADS	MultiCal, Statistical	Matlab + RF Toolbox
Quickwave	Microwave Office	WinCal	
	Ansoft Designer		
	QUCS		

Table 5.1.: Common RF/Microwave Engineering Software sorted Categorically

Almost all of these products are highly refined, professional-grade projects that are extremely sophisticated and powerful. They are also designed for engineers in industry, and embrace a specific work flow. One common trait shared among most of the software packages listed in Tab. 5.1 (Matlab excluded), is an interface-driven design. Because the majority of users interact with software from a graphical user interface (GUI), the accessibility to the core functionalities from a programming interface is generally neglected. While most of the products provide a scripting ability as well as some application programming interface (API), they are mostly clunky and awkward. These characteristics produce software that is very good at accomplishing a specific task, but not good for inventing new applications or interconnecting existing functionality quickly. This severely limits the capabilities of the research scientist.

Another problem, especially relevant for research applications, is the closed-source

nature of products in Tab. 5.1. The academic environment is a principle example of the power and effectiveness of the open-source model, yet many of the fruits of the academic environment, namely the algorithms and analysis techniques, are implemented in a closed-source manner. This impedes scientific progress, and is itself a genuine problem which justifies future study.

Similar drawbacks were recognized nearly 20 years ago at Caltech by David Rutledge and his students. Their solution was to build their own simple, easy-to-use microwave circuit design program called Puff[58]. In the authors own words;

> .. one of the main advantages of Puff over existing CAD programs is its openness. The models used by Puff and the analysis methods are thoroughly documented, and users are invited to contribute to the programs development.

Although antiquated in its interface and software design, Puff has many features, such as it's subnetwork growth algorithm[59] that are worthy incorporation in the next generation of microwave software.

5.2. Current Solutions

There are a few solutions that currently exist which address some of the problems outlined above. The Quiet Universal Circuit Simulator (QUCS)[1] is a modern, open-source RF/microwave circuit simulator written in C++, with impressive functionality and good documentation. However, the architecture of QUCS follows the interface-driven design paradigm, making it an effective tool for the specific purpose that the authors envisioned, but not much else. Since QUCS is open-source, its useful capabilities can be either incorporated into other projects with minimal

---

[1]http://qucs.sourceforge.net/

effort.

Matlab's RF toolbox, on the other hand, provides much of the necessary framework for microwave analysis, such as a n-port network class and cascading/de-embedding operations. Unfortunately, Matlab is not an object-oriented programming language, which limits its scalability, it is closed source, and it is expensive.

In summary, some key problems identified with existing software are interface-driven design, lack of modularity, deficiency of integration, and closed-source philosophy. A flexible, object-oriented, open-source framework for microwave and RF engineering is needed. With community support, such a framework could provide a common foundation for numerous specialized applications such as circuit design, calibration, and network analysis. Many different aspects of microwave engineering could be integrated, providing more creative and powerful solutions than will ever be possible with a set of isolated, specialized tools.

5.3. About `scikit-rf`

`scikit-rf` is a Open-Source, Object-Oriented software package written in the Python programming language[60]. Python was chosen because it is a modern, high-level, cross-platform language with excellent numerical and scientific computing support through the `matplotlib`, `scipy`, and `numpy` modules[48, 61, 62]. Thanks to these foundational libraries, there has recently been an explosion of specialized modules for applications such as statistical analysis[63], machine learning[64], image processing[65], and many others[66].

`scikit-rf` has been made available on-line since August 11, 2009, hosted on Google Code[67] originally, and now on Github[68]. At the time of this writing, the scikit-rf homepage has received over 1,200 visitors from 71 different countries. It has been

designed to provide a general set of objects without constraints, so that the user can build new applications quickly. A brief list of the current capabilities is given below,

- provides n-port network class

- read/write touchstone (.sNp) files

- convert between s, t, y, and z-parameters

- connect n-port networks, and de-embed 2-port networks

- provides basic algebraic operations on networks' scattering parameters

- plot relevant data (dB, Phase, Smith chart)

- 1-port calibration, given any number of standards (least squares)

- 2-port calibration with support for switch-terms.

- provides network set class, used for automating with uncertainty analysis

- can be used with pyvisa for instrument control of some VNA's (partial support for HP8510, HP8720, and R&S ZVA40)

- provides some basic transmission line models (rectangular waveguide, cpw, freespace)

The functionality listed above is throughly documented within a 100+ page user manual, which can be accessed on-line, or downloaded from the project's website, www.scikit-rf.org. Like all Python modules, `scikit-rf` can be used interactively from the command line, used in scripts, or integrated into larger software applications all equally well. The following section demonstrates the unique abilities of `scikit-rf`, by illustrating how to implement the concepts presented in this thesis.

5.4. `scikit-rf` in this thesis

This section illustrates how `scikit-rf` (called `skrf` from within python) was used to implement some of the techniques presented in this thesis. It does not describe the internals of `scikit-rf`, which can be found in the reference section of the user manual or in the source code itself. The first example (sec. 5.4.1) shows how to create a simple one-port calibration, and correct a raw measurement. Building upon this, the second example (sec. 5.4.2) implements the multi-frequency self-calibration technique described in chapter 3, for a SDDL calibration routine. This script also demonstrates a way to visualize the parameter space of the self calibration, as was shown in Fig. 3.1. The final example demonstrates usage the experimentally-based calibration uncertainty estimation presented in chapter 4, for 750-1100GHz (WR-1.0) rectangular waveguide. The uncertainty in the radiating open is measured and compared with theory.

5.4.1. Calibration

Calibrations are performed through the `Calibration` object, which requires at least two pieces of information:

- a list of measured `Networks`

- a list of ideal `Networks`

Optionally, other information can be provided when necessary, such as the type of calibration algorithm, whether the error network should be made reciprocal, etc. When this information is not provided `skrf` will try to determine it through inspection, or use a default value. The default one-port and two-port calibration algorithms are generic in that it will work with any set of standards. If more than

the minimal number of standards is provided, `skrf` will implement a least-squares solution.

```python
import skrf as rf
from pylab import *

# create a Calibration object. This takes as arguments a list of ideal
# and measured  Networks. These can be loaded from touchstone files
cal = rf.Calibration(\
        measured = [\
            rf.Network('measured/short.s1p'),
            rf.Network('measured/delay short 132um.s1p'),
            rf.Network('measured/delay short 85um.s1p'),
            rf.Network('measured/load.s1p'),
            ],
        ideals =[\
            rf.Network('ideals/short.s1p'),
            rf.Network('ideals/delay short 132um.s1p'),
            rf.Network('ideals/delay short 85um.s1p'),
            rf.Network('ideals/load.s1p'),
            ],
        )

# load the raw measurment of a dut from file
dut_meas = rf.Network('dut/radiating open.s1p')
# correct the dut, by applying the calibration
dut_cal = cal.apply_cal(dut_meas)
# load a simulated response for comparison with corrected measurement
dut_sim = rf.Network('simulation/radiating open.s1p')

# plot restults in log magnitude
figure()
title('Return loss of Radiating Open')
dut_cal.plot_s_db(label='Experiment')
dut_sim.plot_s_db(label='Simulated')
draw();show();
```
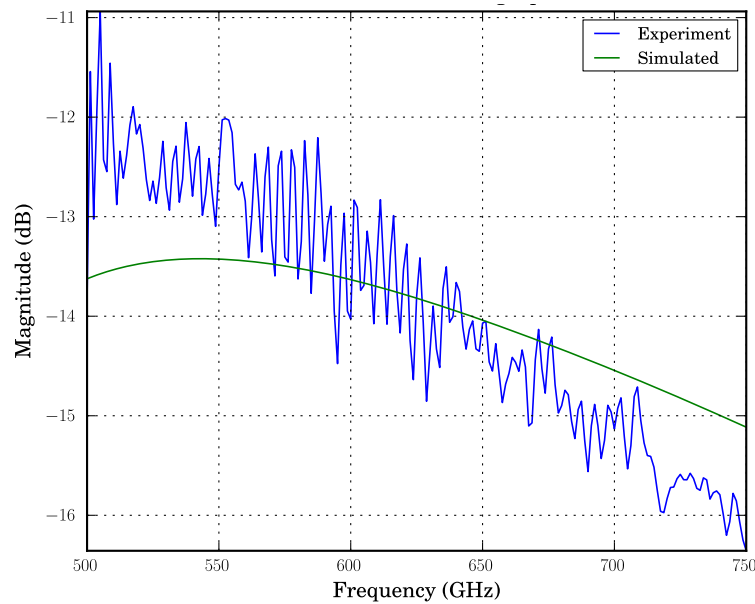
Figure 5.1.: Simulated vs. measured response for a one-port device, calibrated with skrf.

5.4.2. Multi-frequency Self-Calibration

Because of `skrf's` object-oriented approach to calibration, the multi-frequency self-calibration technique described in chapter 3 can be implemented easily. The basic requirements are the ability to modify the parameters of the ideal calibration standards, generate their responses, re-calculate the calibration, and then measure the residual error.

The script below implements a one-port calibration containing a short, two delays of unknown length, and a known load,a calibration set known as SDDL. The self-calibration is accomplished by creating a function, `cal_func`, which takes the delay lengths as an input, runs a calibration, and then returns the mean residual error magnitude. The optimal delay lengths are then found by minimizing the residual error of the calibration. The minimization is accomplished using the `fmin` function of Python module `scipy.optimize`[48], which is an implementation of the Nelder-

Mead simplex algorithm[49].

An exploration of the parameter space is made by sweeping through a grid of different delay short lengths, and calculating the residual error. This two-dimensional space is visualized by a contour plot of the residual error on a log-scale, as shown in Fig. 5.2. From this plot it can be seen that the space appears to be smooth and well-behaved, making minimization plausible.
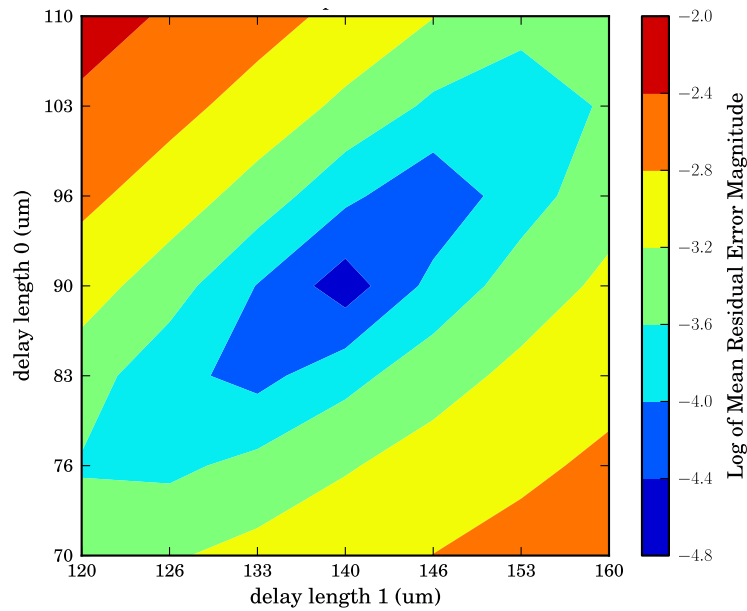
Figure 5.2.: Residual Parameter Space for SDDL Calibration

```python
import skrf as rf
from pylab import *

# measurements
my_measured = [
    rf.Network('measured/short.s1p'),
    rf.Network('measured/delay short 85um.s1p'),
    rf.Network('measured/delay short 132um.s1p'),
    rf.Network('measured/load.s1p'),
    ]

# create a media object for wr-1.5 rectangular waveguide
wg = rf.RectangularWaveguide(
  frequency = my_measured[0].frequency,
  a = 15*rf.mil )


# create a function, which takes the delay short lengths as a tuple,
# runs a calibration, and returns the mean residual error magnitude
def cal_func(d):
  cal = rf.Calibration(\
    ideals = [\
      wg.short(),
      wg.delay_short(d[0]*1e-6),
      wg.delay_short(d[1]*1e-6),
      wg.match(),
      ],
    measured = my_measured,
    )
  cal.run()
  return mean(abs(cal.residuals))


# import fmin from scipy.optimize, a general purpose minimizer
from scipy.optimize import  fmin
# calculate the optimal delay lengths, by minimizing the mean residual
# error of the overdetermined calibration
d_opt = fmin (cal_func, x0 = [80,130])

# create a 2D parameter space for the two delay short lengths and
# calculate the mean residual error at each point.
n = 7 # density of space, nxn
d0_vector =linspace(70,110,n)
d1_vector =linspace(120,160,n)
d_space = array([cal_func([d0,d1]) \
  for d0 in d0_vector \
  for d1 in d1_vector]).reshape(n,n)

# show space as a contour plot
title('Residual Parameter Space for SDDL Calibration')
contourf(log10(d_space))
cb= colorbar()
cb.set_label('Log of Mean Residual Error Magnitude')
grid(0)
yticks(range(n), int_(d0_vector))
xticks(range(n), int_(d1_vector))
ylabel('delay length 0 (um)')
xlabel('delay length 1 (um)')
draw()

show()
```

5.4.3. Calibration Uncertainty Estimation

The experimentally-based technique for calibration uncertainty estimation presented in chapter 4 can be implemented with `skrf` easily. The requirements are an ability to produce a set of `Calibration` objects by combining the different calibration standard measurements, and then use this set correct a single measurement, thereby producing a set of corrected responses. The script below accomplishes this by creating a function, which takes an index parameter and loads measurements that have filenames corresponding to this parameter.

In `skrf`, sets of networks can be represented by a `NetworkSet` object. This object has built-in methods for calculating and plotting common statistical measures on a set of networks. This object is used in the script below to hold the corrected responses of the DUT and plot the results with uncertainty bounds. This experimentally-based calibration uncertainty method has been streamlined in `skrf` with the `CalibrationSet` object, making application and extension of the method very easy. This object was not used in this example in order to more clearly illustrate the concept.

```python
from pylab import *
import skrf as rf
from scipy.constants import *

um = 1e-6

# pull frequency information  from a measurement
frequency = rf.Network('raw/short,0.s1p').frequency
# create a media object for wr-1.0 rectangular waveguide
wg = rf.RectangularWaveguide(frequency, a=10*mil, z0=50)

# function used to generate a Calibration objects
# for a given indecie (k), return a Calibration object for the (kth) of
# set of calibration standard measuremnts. Note this relies on using a
# consistent naming convention on the touchstone files
def cal_func(k):
    cal = rf.Calibration(
        ideals = [
            wg.short(name='short'),
            wg.delay_short(106*um, name='qw'),
            wg.match(name='load'),
            ],
        measured = [
            rf.Network('raw/short,%i.s1p'%k),
            rf.Network('raw/qw,%i.s1p'%k),
            rf.Network('raw/load,%i.s1p'%k),
            ],
        )
    return cal

# create a list of Calibration objects for all measurement sets
cal_list = [cal_func(k) for k in range(3)]

# load the measured dut
dut_meas = rf.Network('dut/ro,0.s1p')
# create a set of corrected dut's by applying the each calibration in cal_list
dut_set = rf.NetworkSet([k.apply_cal(dut_meas) for k in cal_list])
# load the simulated response of dut
dut_sim = rf.Network('sim/ro_hfss.s1p', name = 'ro, HFSS')

# plot simulated vs. measured results with calibraiton uncertainty bounds
figure()
title('WR-1.0 Radiating Open, Magnitude')
dut_set.plot_uncertainty_bounds_s_db(
    n_deviations = 3,
    label = 'Measured $\pm 3\sigma$')
dut_sim.plot_s_db(label='HFSS')
ylim((-20,-10))

figure()
title('WR-1.0 Radiating Open, Phase')
dut_set.plot_uncertainty_bounds_s_deg(
    n_deviations = 3,
    label = 'Measured $\pm 3\sigma$')
dut_sim.plot_s_deg(label='HFSS')
ylim((-120,-60))
draw()
show()
```
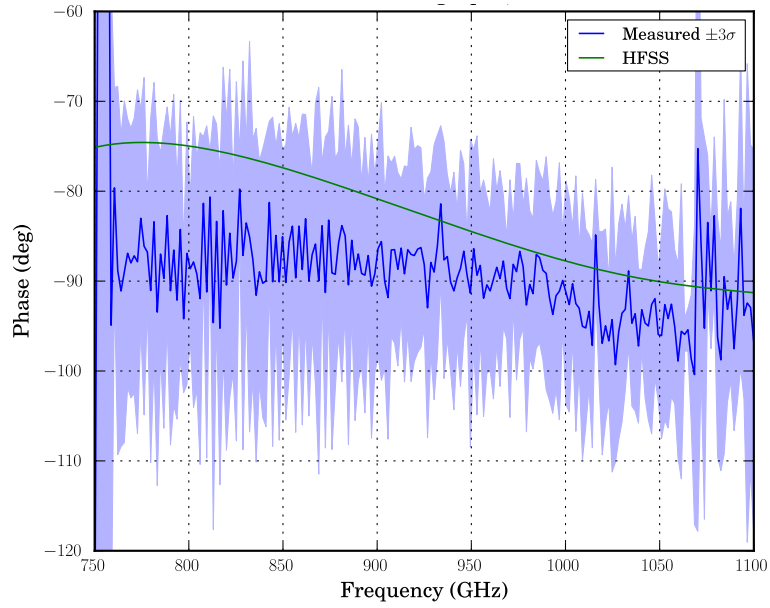
Figure 5.3.: Simulated vs measured return loss phase one-port device with uncertainty bounds, as calculated from experimentally based calibration uncertainty.
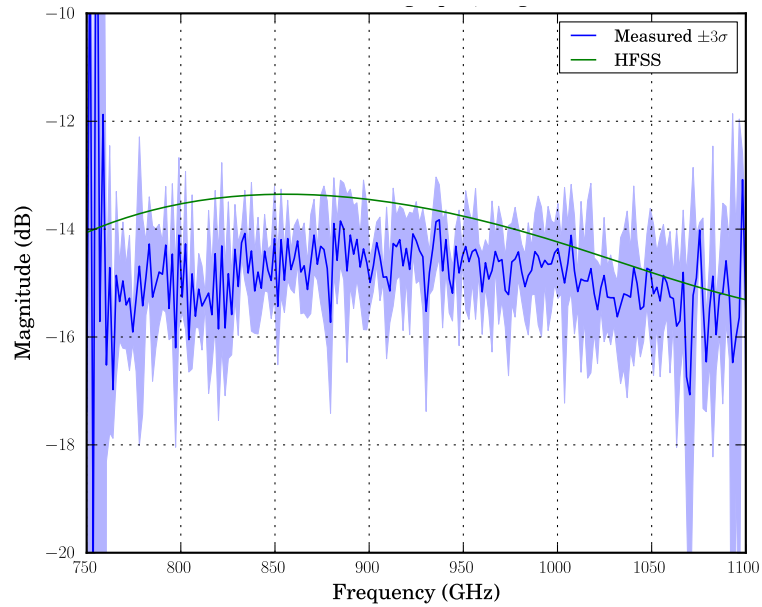
Figure 5.4.: Simulated vs measured return loss magnitude for a one-port device with uncertainty bounds, as calculated from experimentally based calibration uncertainty.

5.5. Conclusion

A series of deficiencies in the existing software tools available to the microwave re-
search scientist have been identified. This motivates the creation of `scikit-rf`, an
open-source, object-oriented python module for microwave engineering. `scikit-rf`'s
unique abilities have been demonstrated by implementing the major concepts devel-
oped in this thesis. More thorough documentation on `scikit-rf`'s capabilities and
usage examples can be found on the project's website www.scikit-rf.org.

6. Conclusion

The field of terahertz science and technology is rapidly expanding, making the need for a proper measurement infrastructure essential. Building this infrastructure is a multi-faceted, community-wide effort. Given its history of millimeter-wave and terahertz technology, the University of Virginia is in a unique position to make vital contributions to this effort. This ability is further strengthened by our proximity to Virginia Diodes and the National Radio Astronomy Observatory, both well-known pioneers in the field. In recent years, UVa has made considerable advances to the measurement infrastructure through the creation of novel components[4, 5, 69], evaluation of standards and connection hardware [2, 3, 12], and measurement techniques[70, 71].

The work presented in this thesis contributes new methods of calibration (chapter 3) and uncertainty analysis (chapter 4), subjects which are generally applicable to all microwave research laboratories. Several practical applications of these techniques are given both in body as well as the appendices. The practical use of these techniques was made possible because of software project scikit-rf, which has been described in chapter 5. Areas of future work are given below.

6.1. Future Work

Because each chapter is intended to be modular, they contain specific suggestions for future work within themselves. Both the multi-frequency self-calibration chapter 3, and the experimentally based uncertainty estimation method of chapter 4 can be extended to a variety of different measurement scenarios, depending upon the community's needs. Given the inability to predict which specific applications will attract the most attention, a more useful endevour is to increase awareness of such techniques, making the software more accessible so that other can use it more easily. The following sections identify general areas of future work which the author plans to pursue, all things willing.

6.1.1. Coherent Approach to Identifying Measurement Limitations

One major problem for the terahertz metrology community is the lack of a coherent and consistent approach in characterizing measurement uncertainty. Such an approach is needed so that the dominant sources of uncertainty can be identified and dealt with. Furthermore, each group has their own measurement and analysis techniques, making it very difficult to draw definitive conclusions. It is therefore recommended that guidelines for uncertainty estimation measurement procedures be explicitly written out, and that raw data be shared openly amongst everyone. This will allow for everyone to share on another's measurement efforts, and all analysis techniques can be applied to the same pool of data.

6.1.2. Scikit-rf

The unique abilities provided by `scikit-rf` have been demonstrated through the applications in this thesis. The level of on-line interest in the project indicates that there is a demand for such functionality. At this point, the major requirement for `scikit-rf` is a developer community. While there are a small number of users that have contributed valuable functionality and bug reports to `scikit-rf`, a dedicated team with sufficient knowledge of the module is needed. Creating a website, moving the code-base to Github, and re-branding the project a *scikit* were all intentional efforts to attract developers, and lower the barrier of contributing code. These efforts have been moderately successful.

It appears that one major limitation to the development of `scikit-rf` is the rarity of the skill-set needed to contribute. There are many python programmers and many microwave engineers, but few python programming microwave engineers. At this point it is unclear whether `scikit-rf` will gain traction and grow, or remain a small project used for a select few.

6.1.3. Geometric Algebra

While studying the Möbius Transformation and non-euclidean geometry, it became clear that the mathematical foundations of electrical engineering, and all of engineering for that matter, could be much improved. This observation is due mainly to the insightful writings of David Hestenes[72], the chief architect and proponent of the modern Geometric Algebra (GA). From an engineers perspective, Geometric Algebra can be most quickly described as the generalization of complex numbers to arbitrary dimension, but it is much more fundamental than that. It is a revolutionary approach to mathematics originally developed in 1800's. Since then, specific

instances of Geometric Algebras have been re-discovered several times, due to the scientific community's lack of awareness. Two examples of such Geometric Algebras are the Pauli and Dirac matrices, invented by the physicists Wolfgang Pauli and Paul Dirac.

In his doctoral thesis, Hestenes showed how to replace the Pauli and Dirac matrices with geometrically interpretable vectors and Spinors, uniting both algebras under a single *Space-time Algebra*[31]. This example can serve as a model for the modernization of electrical network theory. Network theory is the study of networks based on their electrical properties, and is traditionally expressed in terms of complex vector spaces. Measured quantities, such as voltages and currents, are represented by complex vectors, while the networks are represented as complex matrices. Physically meaningful characteristics such as passivity, reciprocity, and symmetry are realized as special conditions of the complex matrices.

The major drawback to this representation is the lack of geometrical interpretation; complex matrices can not be visualized. This leads to difficulties in building an intuitive understanding of what special conditions imply mathematically. Relationships between different basis sets, such as scattering and impedance parameters, are related algebraically but the corresponding geometrical relation is usually not mentioned. Other important quantities, such as complex power are also geometrically intractable.

Another problem is the inefficiency of the algebra. Using complex number inherently singles out a series of two dimensional planes without an easy way to relate other spaces of different dimensions. These are not merely a pedagogical issues, they are severe functional problems, symptomatic of an improper mathematical design[72].

In contrast with the traditional mathematical framework, an alternative approach may be taken by representing the measured quantities, such as voltage, current,

etc, as real vectors. With this construction, the network representations become geometrically interpretable transformations, known as Spinors, which act on the vectors. Special conditions can then be realized as geometrical transformations, such as euclidean and non-euclidean rotations. Such a formulation would allow for the seamless application of projective and non-euclidean geometry to microwave theory, the beginning of which is described in Appendix B.

The key concept to this theory is that the dual role of numbers, the *operational* and the *quantitative*, is perfectly mirrored by the physical reality of network theory. For example, a DUT is an entity which we are trying to measure quantitatively, while an embedding network is thought of as performing an operation. In both cases, the physical entities are the same and their interpretation is created by us, and for our own benefit. The importance of this dichotomy is somewhat subtle, but becomes clear through the use GA. GA not only supports this duality, but provides mechanisms for translation between the two interpretations.

Creating a Geometric Network Theory is only one piece of the much larger project of reformulating the entire electrical engineering curriculum in the modern language of Geometric Algebra. This is no doubt a monumentus task, but it is a necessary one. I realize that this is not a commonly held perspective, and may be perceived as extreme or naive, but I submit that anyone who learns what Geometric Algebra is will become equally zealous about its proliferation.

Acknowledgments

I would like to acknowledge a few of the many people who have made this effort possible.

First, I would like to thank my parents for encouraging me to study engineering and supporting me in my undergraduate studies. Secondly, I am grateful to Chris Welch for introducing me to the power of science as well as providing guidance when it was most needed. And finally, I thank my advisor, Bobby Weikle, for infecting me with his enthusiasm.

A. Individual Studies

A.1. Multi-Reflect One-port Calibration

A.1.1. Introduction

Recently there has been a renewed interest in one-port calibration techniques for mm-wave applications. Achieving high quality measurements using industry trusted two port calibration algorithms such as TRL is proving difficult due to the large amounts of two-port phase drift, and inability to realize line standards in the waveguide medium.

One of the main advantages of TRL is its ability to solve for relevant unknowns of the calibration standards themselves during calibration. Extending this ability to one-port calibrations is desirable, as it will provide the same abilities of TRL for mm-wave measurements in which high quality two-port calibrations are becoming impractical. This concept has been presented before in [**?**]. The treatment given here is slightly different in derivation, making it more consistent with the rest of this thesis, and suggestions toward a closed form solution are given.

A.1.2. Theory

A one-port VNA is to be calibrated using a multiple delayed terminations. Assuming that the termination is unchanging, the only parameter changing from measurement to measurement is the length of the delay line. If enough measurements are made then both the propagation constant of the delay line and reflection coefficient of the termination can be found, along with the parameters of the error network.

Given a linear two-port error network, the measured and actual reflection coefficient

are related by the familiar transforms,

$$\mathbf{a} = \frac{\mathbf{m} - e_{00}}{e_{11}\mathbf{m} - \Delta e} \qquad \mathbf{m} = \frac{\Delta e\,\mathbf{a} - e_{00}}{e_{11}\mathbf{a} - 1} \tag{A.1}$$

The idea is to start with the known reflection coefficient, and remove dependence on all the unknown parameters. The actual complex reflection coefficient of a delayed termination is given by,

$$\mathbf{a} = \Gamma_0 e^{-2j\gamma l} \tag{A.2}$$

Where (Γ_0) is the reflection coefficient at the termination, (γ) is the complex propagation constant and l is the length to the line. To remove the dependence on (Γ_0), take the ratio of (A.9) for two different delay lengths,

$$\frac{\mathbf{a_i}}{\mathbf{a_j}} = e^{-2j\gamma(l_i - l_j)} \tag{A.3}$$

To remove the dependence on (γ), take the natural log of (A.3) and then ratio two different pairs of measurements,

$$\ln\left(\frac{\mathbf{a_i}}{\mathbf{a_j}}\right) = -2j\gamma\,(l_i - l_j) \tag{A.4}$$

$$\frac{\ln\left(\frac{\mathbf{a_i}}{\mathbf{a_j}}\right)}{\ln\left(\frac{\mathbf{a_j}}{\mathbf{a_k}}\right)} = \frac{l_i - l_j}{l_j - l_k} \tag{A.5}$$

Alternatively, this can be written as,

$$\frac{\ln\left(\mathbf{a_i}\right) - \ln\left(\mathbf{a_j}\right)}{\ln\left(\mathbf{a_j}\right) - \ln\left(\mathbf{a_k}\right)} = \frac{l_i - l_j}{l_j - l_k} \tag{A.6}$$

Which provides a nice geometric interpretation that the ratio of distances is equal for delay lengths and log-reflection space.

Note that the actual complex reflection coefficient in (A.6) is related to the measured reflection coefficient by a non-linear function (eq (A.1)), parameterized by the error network scattering parameter matrix (E). Thus, eq (A.6) is non-linear in terms of the desired parameters (e_{ij}), and will therefore require a non-linear least squares solution. Something similar to this is done in [?].

Given that a solution can be found, the propagation constant and the reflection coefficient of the termination are directly calculated by

$$\gamma = \frac{\ln\left(\mathbf{a_i}\right) - \ln\left(\mathbf{a_j}\right)}{2j\left(l_j - l_i\right)} \tag{A.7}$$

$$\Gamma_0 = \frac{\mathbf{a_i}}{e^{-2j\gamma l_i}} \tag{A.8}$$

Now, to count the unknowns. Three complex measurements provide one complex *restraint* equation given by (A.6). Any additional measurement will provide another equation, so the total number of equations is $N - 2$, where N is the number of measurements. Therefore, given that there are three complex unknowns representing the embedding network, we require five total measurements. This is sensible because

we are solving for three complex unknowns of the error network, the propagation constant, and the reflection coefficient, which totals five complex unknowns.

A.1.3. Criticisms

One justified criticism of the technique presented above may be that, in the end, a non-linear least square solution is used, so why not simply employ it from the start, by way of the following measurement equation:

$$\mathbf{m_i} = \frac{\Delta e\, \Gamma_0 e^{-2j\gamma l_i} - e_{00}}{e_{11}\Gamma_0 e^{-2j\gamma l_i} - 1} \tag{A.9}$$

Thus, the advantage of this method are more conceptual than otherwise. However, if the above formulation yielded a closed form solution this would be very advantageous. The following sections document attempts at this.

A.1.4. Unfinished Attempts at Closed Form Solution

A.1.4.1. Polynomial

To arrive at a closed from solution, or at least an easily invertible matrix, we would like to construct a polynomial in terms of the unknowns. Starting with A.5, and labeling the ratio on the right hand side, (l_{ijk})

$$\frac{\ln\left(\frac{\mathbf{a_i}}{\mathbf{a_j}}\right)}{\ln\left(\frac{\mathbf{a_j}}{\mathbf{a_k}}\right)} = \underbrace{\frac{l_i - l_j}{l_j - l_k}}_{l_{ijk}} \tag{A.10}$$

$$\ln\left(\frac{\mathbf{a_i}}{\mathbf{a_j}}\right) = \ln\left(\frac{\mathbf{a_j}}{\mathbf{a_k}}\right) l_{ijk} \tag{A.11}$$

Taking the exponential of both sides and cross multiplying,

$$\frac{\mathbf{a_i}}{\mathbf{a_j}} = \left(\frac{\mathbf{a_j}}{\mathbf{a_k}}\right)^{l_{ijk}} \tag{A.12}$$

$$\mathbf{a_i}\mathbf{a_k}^{\mathbf{l_{ijk}}} = \mathbf{a_j}^{\left(\mathbf{1+l_{ijk}}\right)} \tag{A.13}$$

Now plugging in eq A.1,

$$(m_i - e_{00})(m_k - e_{00})^{l_{ijk}}(m_j e_{11} - \Delta e)^{\left(1+l_{ijk}\right)} = (m_j - e_{00})^{\left(1+l_{ijk}\right)}(m_i e_{11} - \Delta e)(m_k e_{11} - \Delta e)^{l_{ijk}} \tag{A.14}$$

I'm not sure what to do with this. In the special case that all the line length intervals are identical, $l_{ijk} = 1$, and eq A.14 becomes,

$$(m_i - e_{00})(m_k - e_{00})(m_j e_{11} - \Delta e)^2 = (m_j - e_{00})^2(m_i e_{11} - \Delta e)(m_k e_{11} - \Delta e) \tag{A.15}$$

118

This could possibly be expanded and solved for the parameters e_{ij}, or unique relations to them.

A.1.4.2. Matrical

In order to make use of the same linear algebraic methods used for TSD[**?**], and TRL [**?**], one can represent one-port networks as being terminated with a matched through on the far port. Thus it's two-port s-matrix is,

$$S = \begin{bmatrix} s_{11} & 1 \\ 1 & 0 \end{bmatrix} \qquad (A.16)$$

The scattering transfer parameters for a one-port is,

$$T = \begin{bmatrix} 1 & s_{11} \\ 0 & 1 \end{bmatrix} \qquad (A.17)$$

The measurement equation for a delayed termination written in terms of wave cascading matrices is thus

$$M = E \cdot L \cdot G \qquad (A.18)$$

For the ith delay line this becomes,

$$M_i \;=\; E \cdot L_i \cdot G \tag{A.19}$$

$$\begin{bmatrix} 1 & m_i \\ 0 & 1 \end{bmatrix} \;=\; \frac{1}{e_{21}} \begin{bmatrix} -\Delta e & e_{11} \\ -e_{22} & 1 \end{bmatrix} \cdot \begin{bmatrix} e^{-j\gamma l_i} & 0 \\ 0 & e^{j\gamma l_i} \end{bmatrix} \cdot \begin{bmatrix} 1 & \Gamma_0 \\ 0 & 1 \end{bmatrix} \tag{A.20}$$

Where M is the measurement, E is the error network, L is the delay line, and G is the termination. The fact that neither E or G are changing from measurement to measurement, can probably be used to create a closed form expression for E.

A.2. Artifacts in Two-tier Calibration

While measuring some two-port devices using the tiered calibration scheme described in sec. 2.2.2, a curious artifact was noticed. The artifact was that the de-embedded two-port DUT's look more asymmetrical than expected. Specifically, the DUT's reflection on the test-port side, consistently showed much smaller reflection magnitude. A plot of calibrated s-parameters for two straight sections of waveguide are shown in Fig. A.1, illustrating the effect. In this plot, port 1 is the VNA test-port and port 2 is the far port.
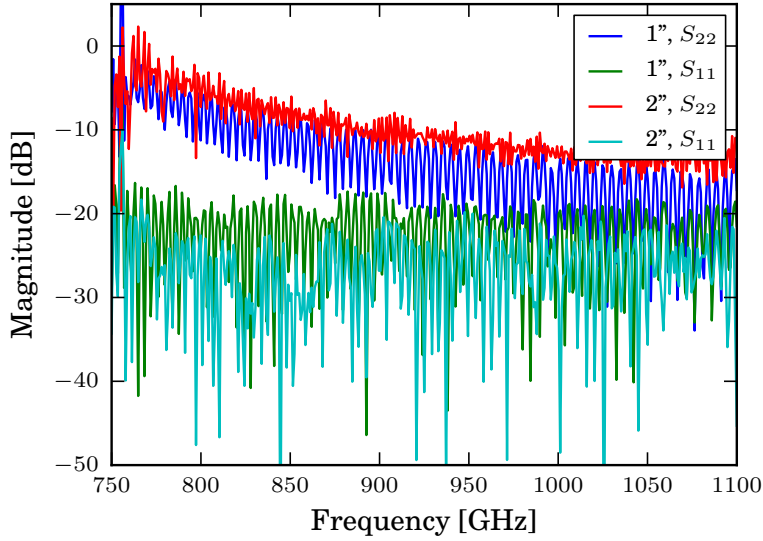
Figure A.1.: Corrected responses of straight sections of rectangular waveguides, illustrating the two-tier asymmetry artifact.

From this plot, it can be seen that not only are the DUT's asymmetrical, but they are reproducibly asymmetrical in a consistent way. To confirm that this is indeed an artifact and not a coincidence, we took a single DUT (2″ section of waveguide) and calibrated it in both possible orientations, meaning we swapped ports 1 and 2. The results were that the test-port side of the corrected DUT exhibited a smaller reflection magnitude than the far side, regardless of the DUT's orientation.

After much fruitless reasoning, a synthetic calibration was created to simulate the entire experiment. It was then discovered, that by inducing error in the calibration standard definitions at each tier, the artifact could be re-created. Specifically, if the delay length of the delay shorts are off by 5-10μm then an amount of asymmetry is produced that is consistent with the measurements.

The results of the synthetic experiment is shown below in Fig. A.2. This plot shows the ideal response of a waveguide section on left, the result from the synthetic experiment in the middle, and the real measurement on the right. The ideal response's transmission coefficient was taken from measurements, so as to most resemble the real measurement, but the reflection at each port was set to 0.
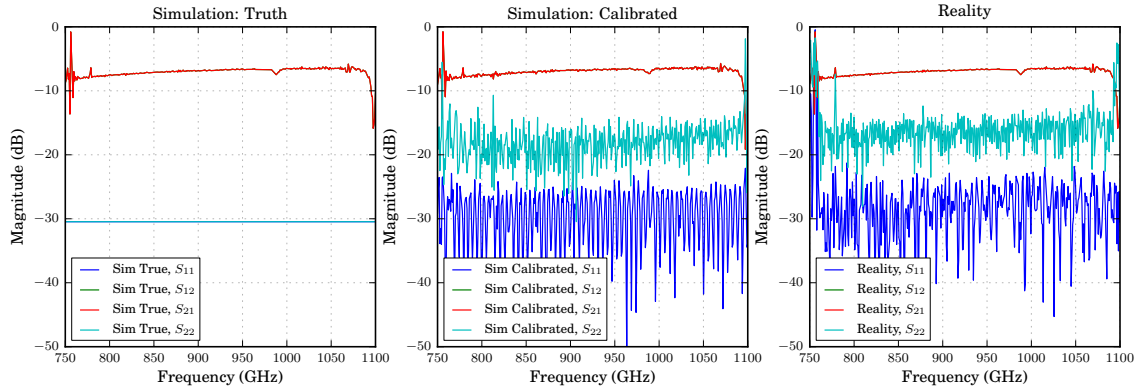
Figure A.2.: Results from the synthetic experiment, illustrating simulated true response, the simulated calibrated response, and the actual measurements.

In conclusion, this artifact does not appear to be inherent limitation of the technique, but rather a manifestation of calibration error. In fact, this effect may be more a blessing than a curse, for it may be possible to use this artifact as a way to estimate the measurement accuracy for this type of calibration. Alternatively, if the DUT of the second tier is known to be symmetric, as is the case here, then only two calibration standards are required to fully determine its s-parameters. This constraint on the embedding network will convert biased errors, such as those created

this artifact, into unbiased errors. This way the unbiased errors can be measured and used incorporated into the uncertainty estimates. This solution has been presented [14], and is discussed below.

A.3. Calibration of a Reciprocal, Symmetric Error Network

In the special-case that the embedding network of a one-port calibration is reciprocal and symmetric, the error network can be determined by measuring only two calibration standards[14]. This is reasonable, and possibly obvious, being that the scattering matrix for this special case only contains two independent quantities. Previously it has been shown that systematic error in the calibration procedure, such as incorrectly defined calibration standards, can produce artifacts of asymmetry in a symmetric DUT. Therefore, a calibration procedure that enforces the symmetry condition would convert these artifacts to measurable residual error, given that the same number of standards are used.

The measured reflection of a two-port error network terminated with a one-port standard is given by:

$$\mathbf{m} = e_{00} + \frac{e_{10}e_{01} \cdot \mathbf{a}}{1 - e_{11}\mathbf{a}} \tag{A.21}$$

This is linearized by,

$$e_{11}\mathbf{m}\mathbf{a} - \Delta e\,\mathbf{a} + e_{00} = \mathbf{m} \tag{A.22}$$

Imposing the symmetry condition on the error network, $(e_{00} = e_{11})$,

$$e_{11}(\mathbf{m}\,\mathbf{a} + \mathbf{1}) - \Delta e\,\mathbf{a} = \mathbf{m} \tag{A.23}$$

Given two measurements $(\mathbf{m}_{1\ldots2})$ of calibration standards with known responses $(\mathbf{a}_{1\ldots2})$, the unknowns $(e_{00}, \Delta e)$ can be directly found using matrix inversion.

$$\underbrace{\begin{bmatrix} (\mathbf{m}_1 \cdot \mathbf{a}_1 + 1) & -\mathbf{a}_1 \\ (\mathbf{m}_2 \cdot \mathbf{a}_2 + 1) & -\mathbf{a}_2 \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} e_{11} \\ \Delta e \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}}_{\mathbf{b}} \tag{A.24}$$

$$\begin{aligned} \mathbf{A} \cdot \mathbf{x} &= \mathbf{b} \\ \mathbf{x} &= \mathbf{A}^{-1} \cdot \mathbf{b} \end{aligned} \tag{A.25}$$

Given additional measurements, (A.24) becomes overdetermined, and the standard linear least squares method can be applied.

A.4. Alternative One-port Calibration Formulation Station

The following is an alternative closed form solution to one port calibration, providing no major benefits that the author knows of. It was derived without an application

in mind, but in hind-site, the following observations can be made. The solution produces a pair of equations linear in (e_{00}) and (e_{11}) only, as opposed to (e_{00}), (e_{11}), and (Δe). Therefore the vector to be solved for has consistent units for each components. Another minor advantage is that the special case of a reciprocal symmetric error network is reduced to division.

The measured reflection of a two-port error network terminated with a one-port standard is given by:

$$\mathbf{m} = e_{00} + \frac{e_{10}e_{01} \cdot \mathbf{a}}{1 - e_{11}\mathbf{a}} \tag{A.26}$$

Looking at the complex ratio between actual and measured reflection of a given standard, and separating terms which do and don't have a dependence on the terminating standard itself.

$$\mathbf{m} - \mathbf{m}\mathbf{a}e_{11} = e_{00} - e_{00}e_{11}\mathbf{a} + e_{10}e_{01}\mathbf{a} \tag{A.27}$$

$$\frac{\mathbf{m}}{\mathbf{a}} = \underbrace{e_{10}e_{01} - e_{00}e_{11}}_{\text{constant across standards}} + \underbrace{\mathbf{m}e_{11} + \frac{e_{00}}{\mathbf{a}}}_{\text{changing ccross standards}} \tag{A.28}$$

It then seems intuitive to difference this complex ratio for two different *measured-actual* standard pairs, to remove the quantities without dependence on the termi-

nating standards. If this is done for two pairs indexed , i and j,

$$\frac{\mathbf{m}_i}{\mathbf{a}_i} - \frac{\mathbf{m_j}}{\mathbf{a}_j} = e_{11} \cdot (\mathbf{m}_i - \mathbf{m}_j) + e_{00} \cdot \left(\frac{1}{\mathbf{a}_i} - \frac{1}{\mathbf{a}_j} \right) \tag{A.29}$$

There are only two unknowns in this equation, e_{11} and e_{00}. These may be solved for if you have two equations, which would require three standards. This makes sense. Making some notation changes to improve readability.

$$\underbrace{\frac{\mathbf{m}_i}{\mathbf{a}_i} - \frac{\mathbf{m}_j}{\mathbf{a}_j}}_{P_{ij}} = e_{11} \cdot \underbrace{(\mathbf{m}_i - \mathbf{m}_j)}_{Q_{ij}} + e_{00} \cdot \underbrace{\left(\frac{1}{\mathbf{a}_i} - \frac{1}{\mathbf{a}_j} \right)}_{R_{ij}}$$

$$P_{ij} = e_{11} \cdot Q_{ij} + e_{00} \cdot R_{ij}$$

Given three standards this system of equations may be solved,

$$\begin{bmatrix} P_{12} \\ P_{23} \end{bmatrix} = \begin{bmatrix} Q_{12} & R_{12} \\ Q_{23} & R_{23} \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{00} \end{bmatrix}$$

The error coefficients are found by

$$\begin{bmatrix} e_{11} \\ e_{00} \end{bmatrix} = \begin{bmatrix} Q_{12} & R_{12} \\ Q_{23} & R_{23} \end{bmatrix}^{-1} \begin{bmatrix} P_{12} \\ P_{23} \end{bmatrix}$$

The reflection tracking term $(e_{10}e_{01})$ can be solved by evaluating equation A.28, given $e_{00},$ and e_{11}.

Special Case: Symmetric Error Network

Examining the special case of a symmetric error network, by taking eq A.29, and enforcing the symmetry condition; $e_{00} = e_{11}$, yields

$$e_{00} = e_{11} \quad = \quad \frac{\frac{\mathbf{m}_i}{\mathbf{a}_i} - \frac{\mathbf{m}_j}{\mathbf{a}_j}}{\left(\mathbf{m}_i - \mathbf{m}_j + \frac{1}{\mathbf{a}_i} - \frac{1}{\mathbf{a}_j}\right)}$$

Using eq A.28, the reflection tracking term can be found by,

$$e_{10}e_{01} = e_{00}^2 - e_{11}\left(\mathbf{m} + \frac{1}{\mathbf{a}}\right) + \frac{\mathbf{m}}{\mathbf{a}}$$

This formulation is convenient because the calibration calculation is reduced to division.

A.5. Uncertainty in Multiple Tiered Calibrations

A.5.1. Summary

In this section, WR-1.5 waveguide twists are measured using a one-port two tier calibration and confidence bounds are estimated. The confidence bounds are created using a multi-tiered application of the experimental uncertainty estimation technique described in (chapter 4). As a diagnosis tool, the error metrics described in (sec. 2.3.3), are employed and successfully detect a systematic calibration error caused by the experimenter.

A.5.2. Introduction

A measurement needed to characterize a WR-1.5 waveguide twist as described in [69]. Using a standard two-port setup is undesirable for two reasons. First, one of the extension heads must be rotated in between calibration and measurement of the twist. This induces calibration error, due to the movement of the cables. The second problem is that when one head is rotated, the alignment pins on the waveguide flanges interfere with on another. To circumvent these problems a one-port two-tiered measurement was made instead. To ensure that higher order modes were not present at the second tier reference plane, a .8" waveguide section was placed on the far port of the twist. Such a setup requires that the .8" waveguide section be characterized as well. This meant that an addition tier was needed.

Because two twists were to be measured, the second tier was calibrated twice, upon replacement of the twist. All tiers are labeled and illustrated in the block diagram shown in Figure Fig. A.3.
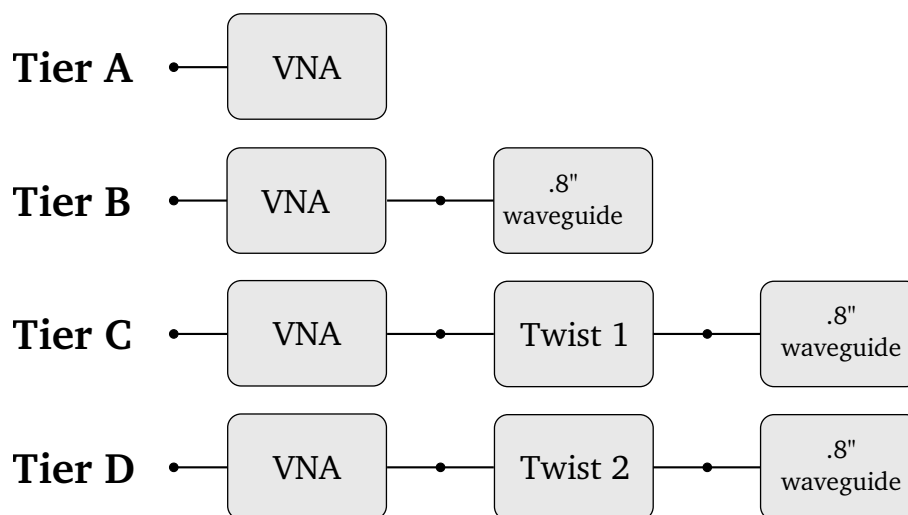
Figure A.3.: Block diagram of calibration tiers

At each tier the following four standards were measured,

1. Flush Short

2. Radiating Open

3. $\frac{\lambda}{4}$-Delay Short (milled type)

4. $\frac{\lambda}{8}$- Delay Short (shim type)

Each standard was measured three times. Between each measurements a given standard was disconnected and re-connected. This was done to capture the uncertainties associated with the flange. Despite evidence suggesting that the shim style delay-short standards are inferior to the milled design[10], the $\frac{\lambda}{8}$-delay short shim was used due to lack of other standards.

A.5.3. Initial Problems

After some insensible results, the error metrics described in (sec. 2.3.3), were employed to measure calibration quality at each tier. The results are shown in Fig. A.4-Fig. A.7. It can be seen that the biased error for calibration tiers 'C' and 'D' are

significantly larger than that for 'A' and 'B'. From experience, it was known that a typical cause for such an error is incorrectly defined calibration standards. It was hypothesized that the $\frac{\lambda}{8}$-delay short shim was not correctly oriented on these tiers, which has been a common problem in the past.

HFSS simulation of a 90deg rotated $\frac{\lambda}{8}$ delay short shim, and talks with Qaing, suggested that this was the case. The measurements of these standards where then removed from the calibration tiers 'C' and 'D', and reasonable results where attained.
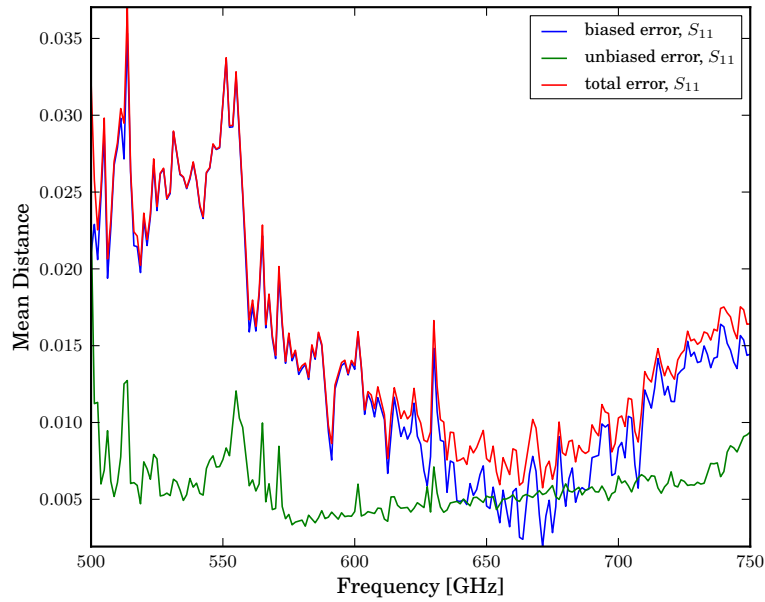
Figure A.4.: Error Metrics for Tier A
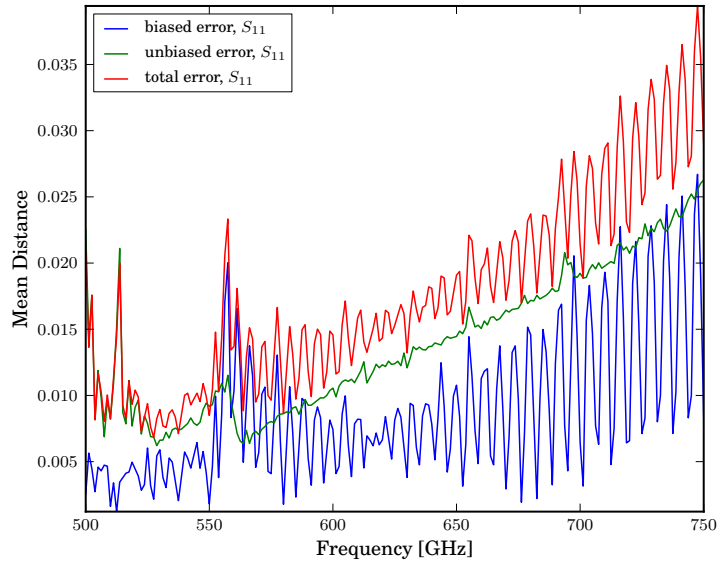
Figure A.5.: Error Metrics for Tier B
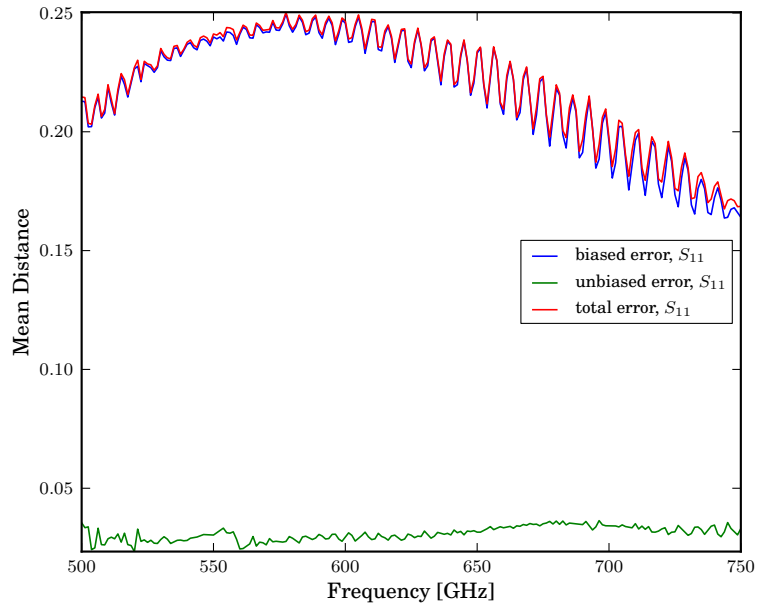
Figure A.6.: Error Metrics for Tier C
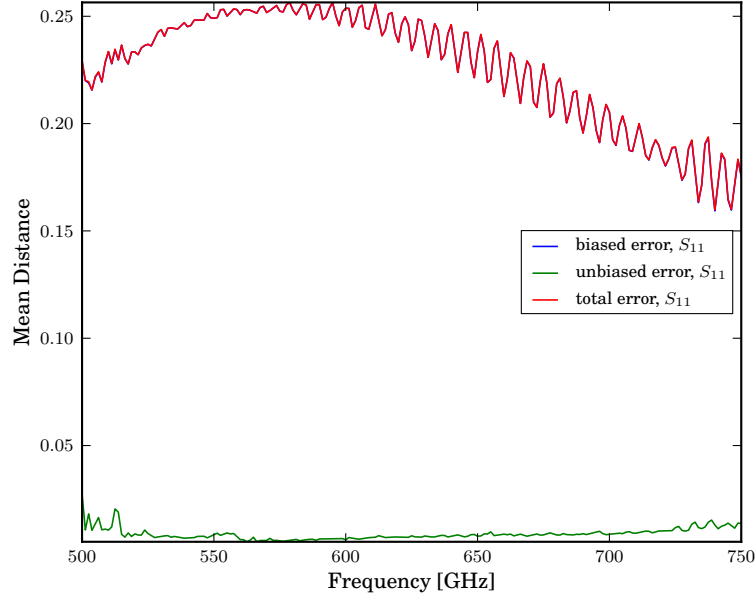
Figure A.7.: Error Metrics for Tier D

A.5.4. Results

The magnitude of the twists's insertions loss with $\pm 3\sigma$ confidence bounds is shown below in Figure Fig. A.8. Plots of real and imaginary or phase are difficult to view due to numerical artifacts from the sign ambiguity caused by de-embedding a reciprocal network.
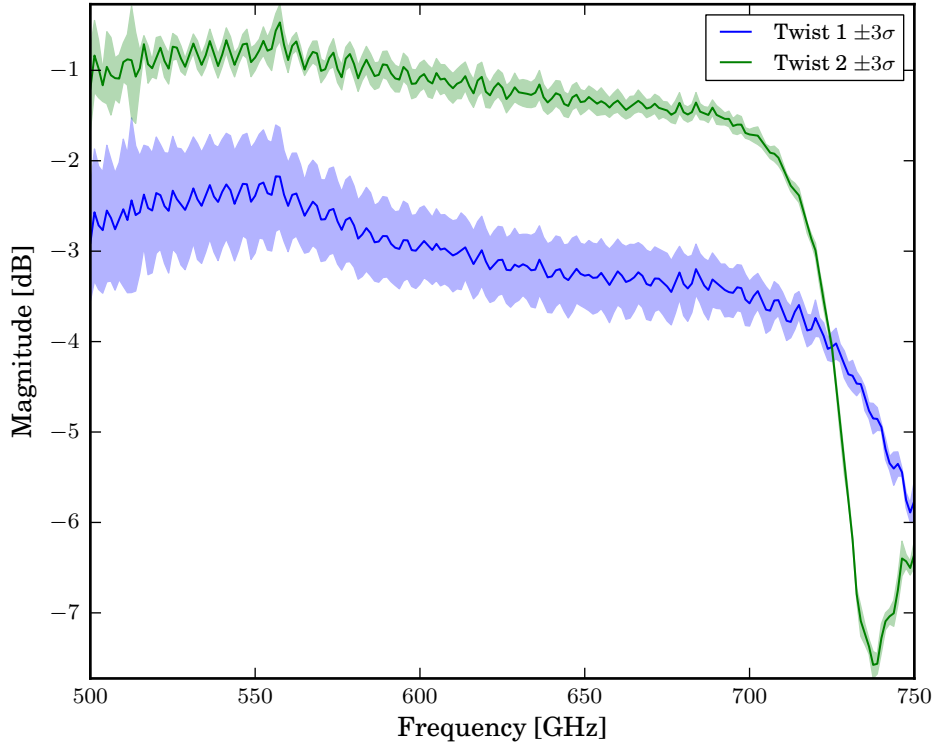
Figure A.8.: Twist's Insertion Loss with Confidence Bounds

The passivity for each port of both twist's are shown in Fig. A.9 (the legend entries are incorrect). Passivity is defined as total power received, from excitation at a single port, given by

$$\text{passivity port 1} = |S_{11}|^2 + |S_{21}|^2$$
$$\text{passivity port 2} = |S_{22}|^2 + |S_{12}|^2$$

This indicates that a major cause of the low insertion loss is due to power dissipation of some sort, perhaps because of non-planarity of the shims.

133
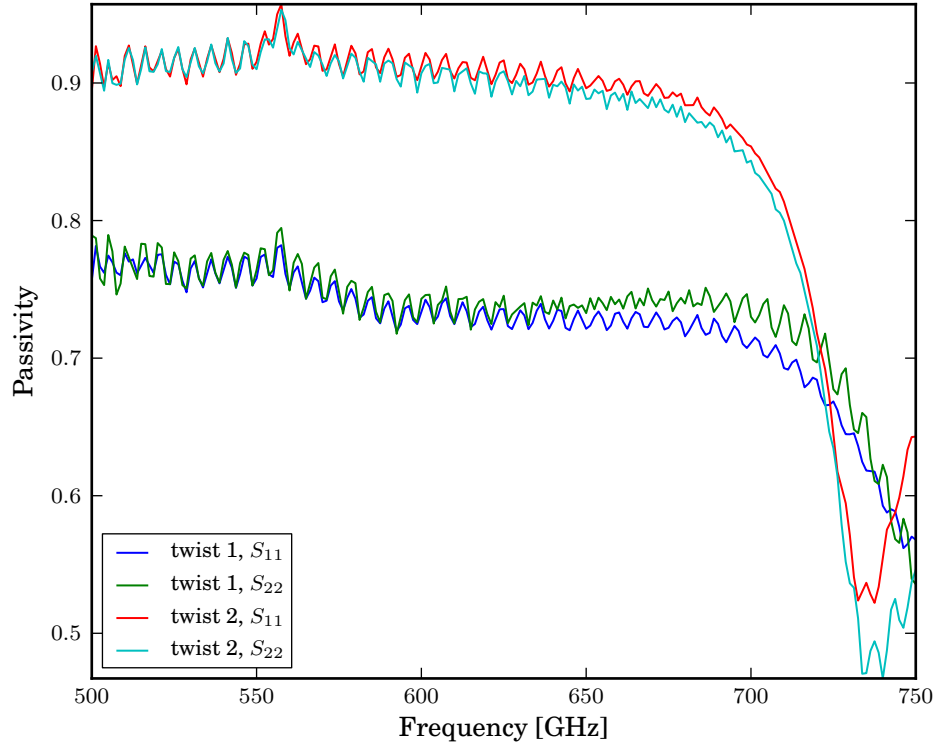
Figure A.9.: Twist's Passivity Metrics

A.5.5. Details of the Multi-tiered Uncertainty Calculation

The uncertainty in the de-embedded DUT's were calculated using the technique described in chapter 4. Because the calibration was multi-tiered, the uncertainty calculations are non-trivial and require explanation. Refer to sec. 2.2.2 for an explanation of tiered calibration. A diagram illustrating the calibration tiers is re-illustrated below in Fig. A.3 with concise notation to make the following mathematics are more readable.
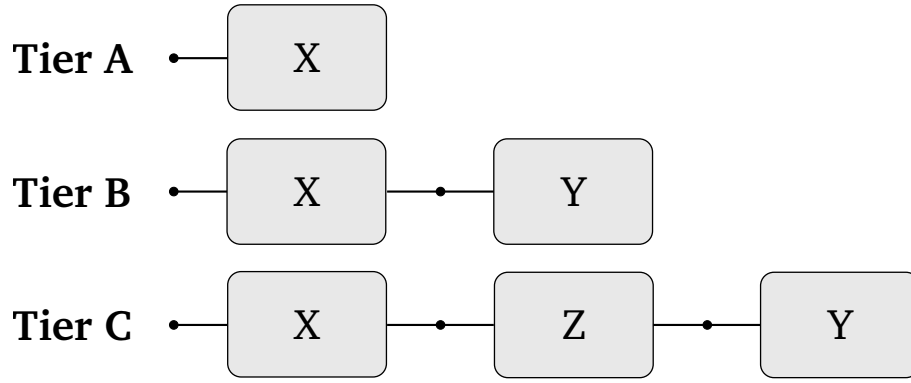
Figure A.10.: Block diagram of calibration tiers

In the digram shown in Fig. A.10, network Z represents the DUT to be characterized. These relations can be written mathematically by,

$$
\begin{aligned}
\text{Tier A:} \quad & \mathbf{X} \\
\text{Tier B:} \quad & \mathbf{XY} \\
\text{Tier C:} \quad & \mathbf{XZY}
\end{aligned}
$$

Once $\mathbf{X}$ is determined, the estimate of the network $\mathbf{Z}$ can then be calculated in two steps,

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{X}^{-1} \cdot (\mathbf{XY}) \\
\mathbf{Z} &= \mathbf{X}^{-1} \cdot (\mathbf{XZY}) \cdot \mathbf{Y}^{-1}
\end{aligned}
$$

Given that there are i-measurement sets made at tier A, a set of networks, $\mathbf{X}_i$, can

be calculated. Likewise, if there are j-measurement sets made at tier B and k at tier C, sets of estimate networks, $(\mathbf{XY})_{\mathbf{j}}$ and $(\mathbf{XZY})_{k}$, can be calculated similarly.

$$\begin{aligned}
&\text{Tier A:} \quad \mathbf{X_i} \\
&\text{Tier B:} \quad (\mathbf{XY})_{\mathbf{j}} \\
&\text{Tier C:} \quad (\mathbf{XZY})_{\mathbf{k}}
\end{aligned}$$

For the purposes of uncertainty estimation it is most straight-forward to calculate the network $\mathbf{Z}$ in a single step so that the size of the the resultant set is clear. The set of $\mathbf{Z}$'s estimated from the sets of each calibration tier is written,

$$\mathbf{Z_{ijk}} = \mathbf{X_i^{-1}} \cdot (\mathbf{XZY})_{\mathbf{k}} \cdot \left(\mathbf{X_i^{-1}} \, (\mathbf{XY})_{\mathbf{j}} \right)^{-\mathbf{1}}$$

The uncertainty in the network $\mathbf{Z}$ can be estimated directly from the set $\mathbf{Z_{ijk}}$. As mentioned in chapter 4, the combinatorial function is chosen to generate the sets is left up to experimenter. For computational speed purposes, I have chosen to generate the individual error networks for each tier using a dot product, and generated the resultant estimate of the $\mathbf{Z}$, by the cartesian product. The sets $\mathbf{X}_i$, $(\mathbf{XY})_{\mathbf{j}}$ and $(\mathbf{XZY})_k$ are each of length 3, and the resultant set $\mathbf{Z_{ijk}}$ is of length 27 $(3 \cdot 3 \cdot 3)$.

A.5.6. Conclusion

Two WR-1.5 rectangular waveguide twists were measured and confidence bounds where estimated, using a multi-tiered experimentally based uncertainty estimation

technique presented in (chapter 4). The error metrics described in sec. 2.3.3 were employed as a way to measure calibration quality at each tier, and successfully reflected that a systematic calibration error was present in two of the four calibrations. It was deduced that this error was produced by the experimenter, and the erroneous data was removed, thereby producing reasonable meaningful results.

A.6. Comparison of Ideal Standard Models on Two-tier Calibration of a Waveguide-to-Coplanar waveguide Probe

A.6.1. Intro

Until now, the second-tier CPW calibrations used to characterize the THz on-wafer probe have been made using simple analytical models for the ideal responses of the delay short lines[1]. The accuracy of these models was challenged by a Matt Bauwens. Therefore, a study was made to determine if using more sophisticated finite-element model for the ideal standards could provide an improvement in calibration accuracy.

Improvement in calibration accuracy is quantified by overall reduction in the average magnitude of residuals, the same measure used the self-calibration routine described in (chapter 3). Corrected measurements of the standards themselves are also presented because they yield more interpretable results. Once the improvement in the calibration accuracy is quantified, the effect of the different ideal definitions on the de-embedded probes s-parameters is calculated.

A.6.2. Explanation of Simple Analytical Models

The analytical models are based on simple TEM transmission line theory. The relative permativity was chosen to produce a propagation constant which agreed with port specifications as found from a HFSS simulation. Only the component responsible for propagation was used (either real of imaginary depending on convention).

---

[1]It should be noted that Theodore Reck was conducting similar calibrations at this time,using more sophisticated analytic models that included conductor loss. Therefore, this comparison should not be applied to his results.

This model neglects both radiative and conductor loss, and assumes a non-changing
propagation constant across the entire band.

A.6.3. Explanation of HFSS Model

The geometry of the model created in HFSS is shown below in (Fig. A.11).

The simulated reflection coefficients of the delay short standards in shown in (Fig. A.12).
L1 is a flush CPW short, generated by de-embedding the model shown in (Fig. A.11)
up to the termination. The other lines (L2-L5) are identical terminations at progres-
sively longer delays, evenly distributed between 0 and 180° , the endpoints excluded.
The effect of the short termination can be seen in the response of L1. The radiation
and conductor loss are shown by the spiraling inwardness of the successively longer
delays.
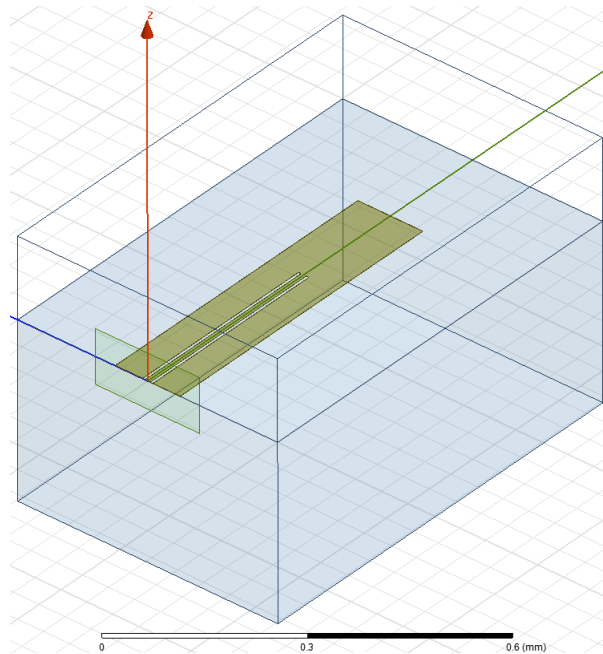
Figure A.11.: HFSS model of CPW delayed short.
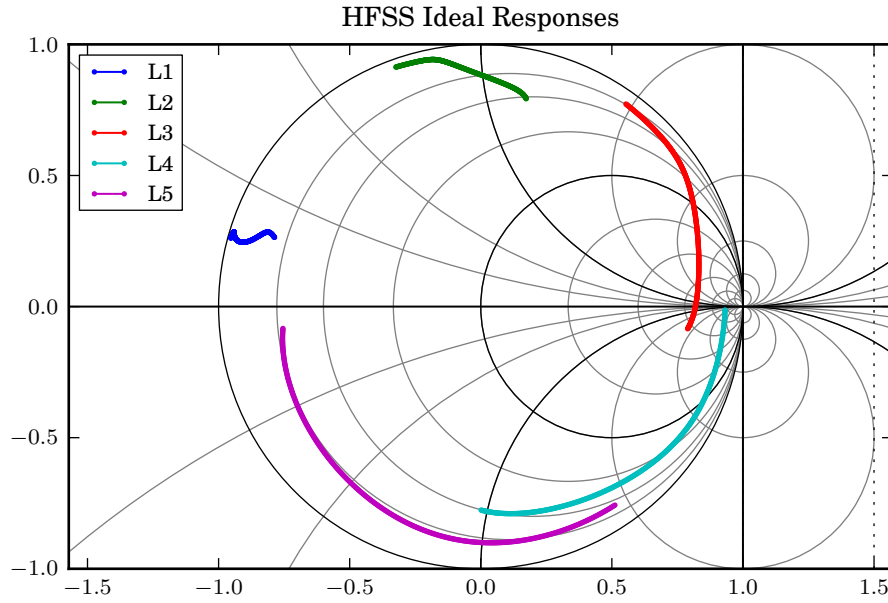
Figure A.12.: Simulated responses of the CPW delay shorts, generated using
HFSS.

A.6.4. Mean Residual Error

Two calibrations were run using the same measurement. One calibration used the
HFSS full-wave models for the ideal standards, and the other used simple analytical
models as described above. The average residual magnitudes for each calibration vs
frequency are shown in (Fig. A.13). From this figure it can be seen that the HFSS
models reduce the mean residual error by about an order of magnitude across most
of the band.
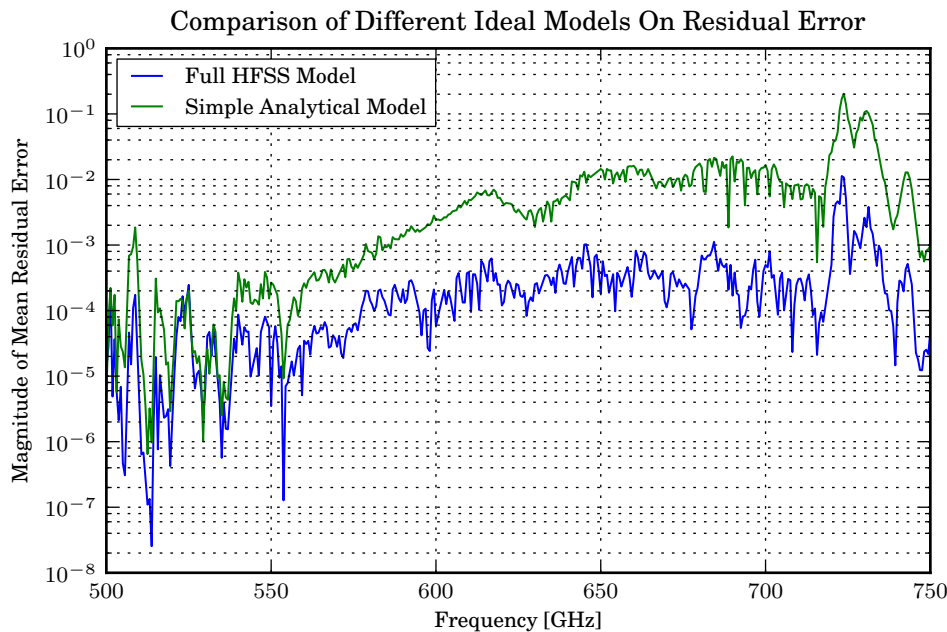
Figure A.13.: Comparison of average residual magnitudes for two different calibrations.

A.6.5. Effect on De-embed Probe's S-parameters

Although improving calibration accuracy is a good thing in and of itself, the more interesting measurement at the time of this study was the change in the de-embedded probe's s-parameters. Two different estimates for the probe's s-parameters are shown in (Fig. A.14) and (Fig. A.15). These estimates were calculated using the same first tier calibration, but different second tier calibrations. The second tier calibration differed by their ideal standard definitions, as described in the previous section.

The difference in the insertion loss is approximately 1dB over the upper 2/3 of the band, with the HFSS models estimating a less loss in the probe itself. The estimated return losses are significantly different on the CPW side of probe only. The difference appears that the HFSS models have a lower return loss in the upper half of the band by 2dB or so.

141
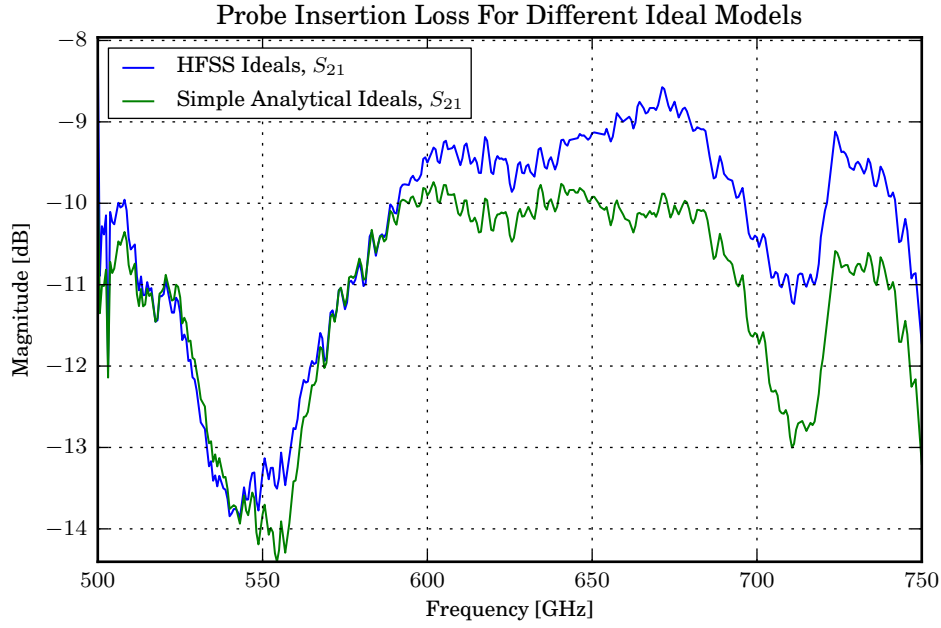
Figure A.14.: De-embedded probes insertion loss, as determined from two different
ideal standard definitions on the second tier calibration.
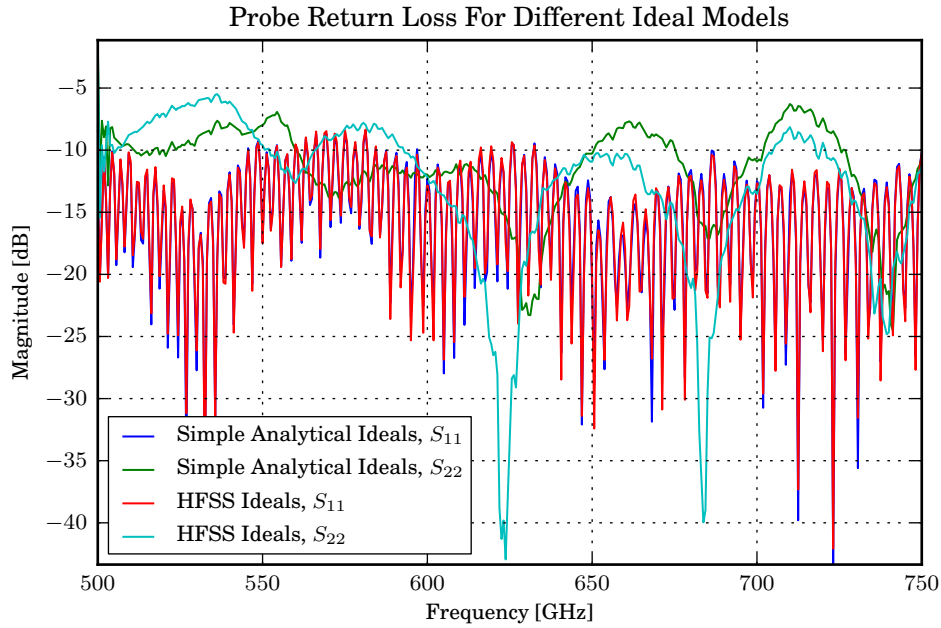
Figure A.15.: De-embedded probes return loss, as determined from two different
ideal standard definitions on the second tier calibration.

A.6.6. Conclusion

The HFSS models produce the following improvements over the simple analytical
models:

- improvement in the magnitude of the mean residual error by approximately
  and order of magnitude for the entire band.

- triple the de-trended phase accuracy on verification delay shorts.

- approximately 1dB difference in probe response for about 2/3 of the band.

Verification measurements of the CPW calibration standards are shown below.
These are provided for those who have experience with similar measurements, but
find the mean residual error a hard quantity to interpret (myself included!).

A.6.7. Verification Measurements

A.6.7.1. Delay Short Magnitude
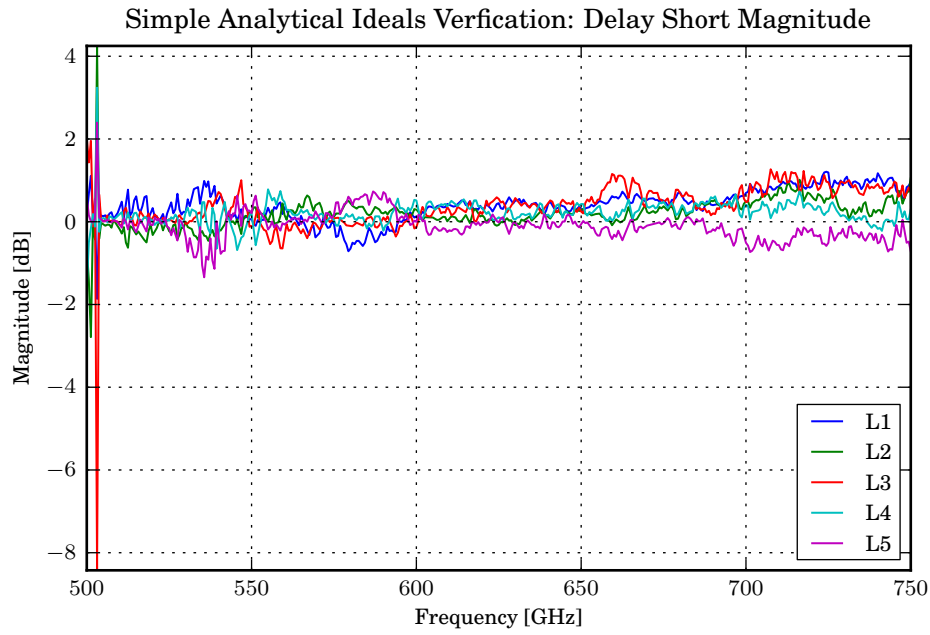
Figure A.16.: (Ideals Response = 0 (duh))
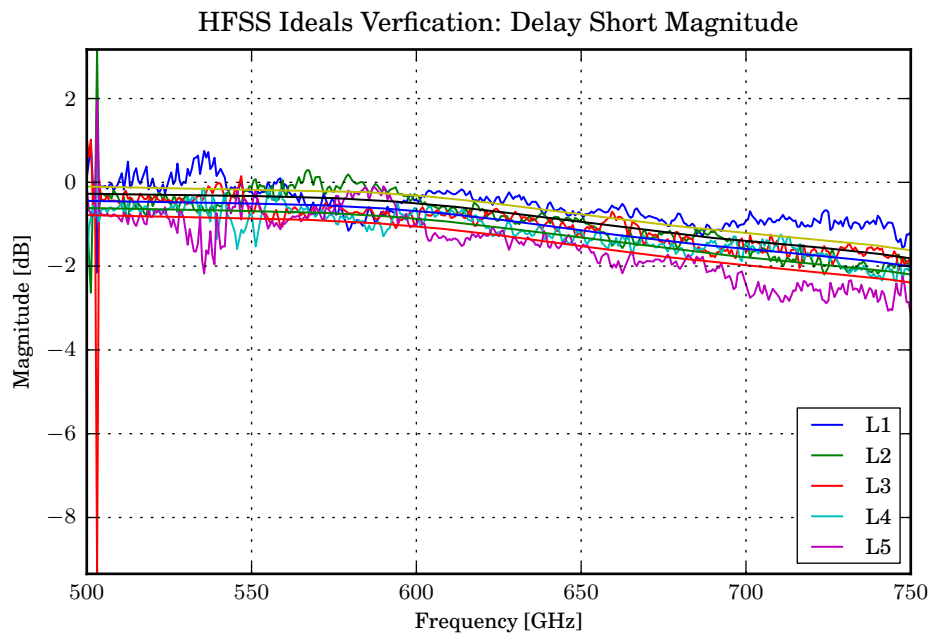
Figure A.17.: (Ideal Responses Shown on graph)

A.6.7.2. Delay Short Phase
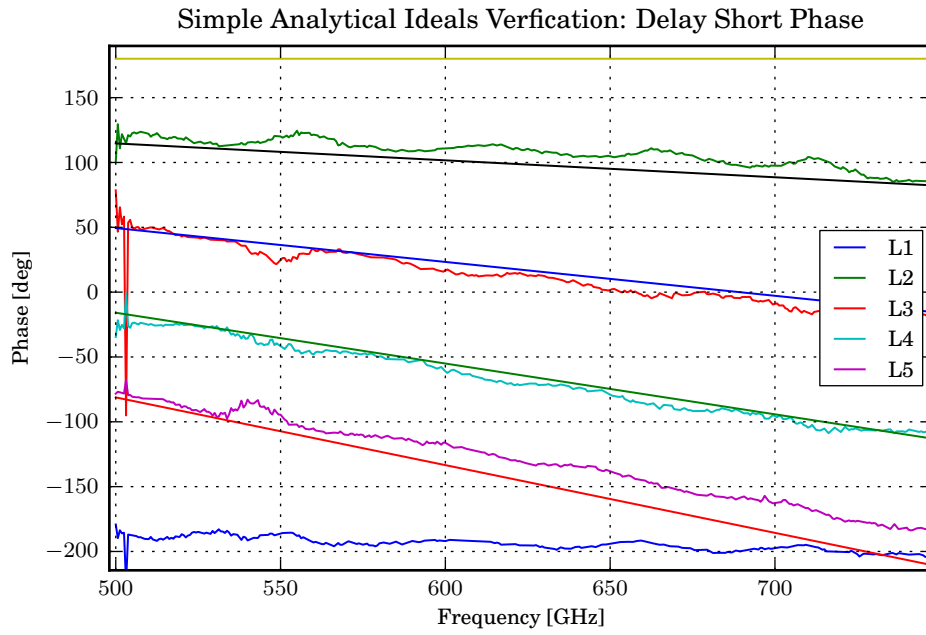
Figure A.18.:
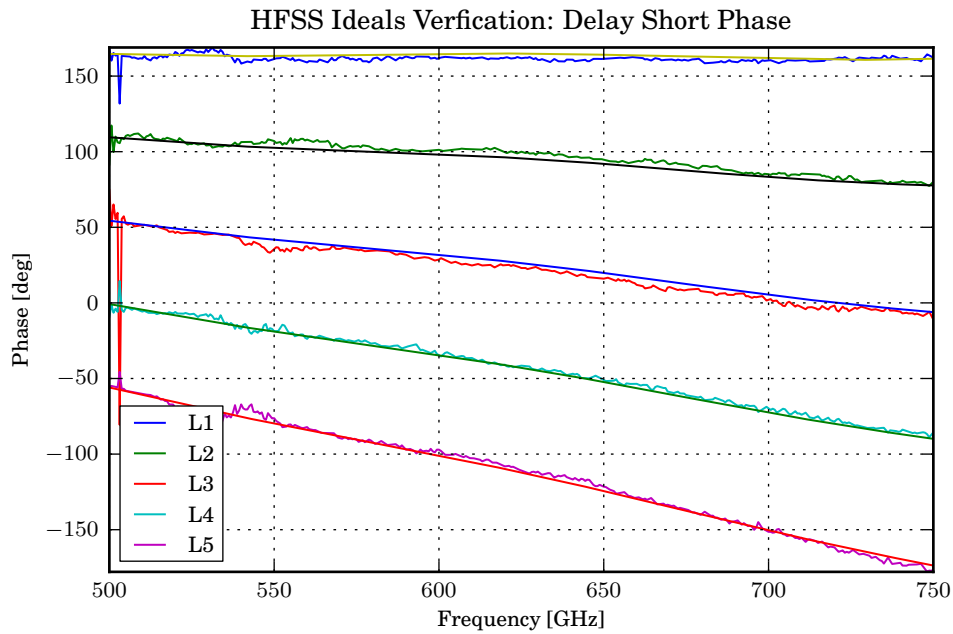
Figure A.19.:

A.6.7.3. De-trended Delay Short Phase

A more quantitative way to view the phase response of verification standards is
through 'de-trended' phase. This is the phase difference between the ideal response,
and the calibrated response.
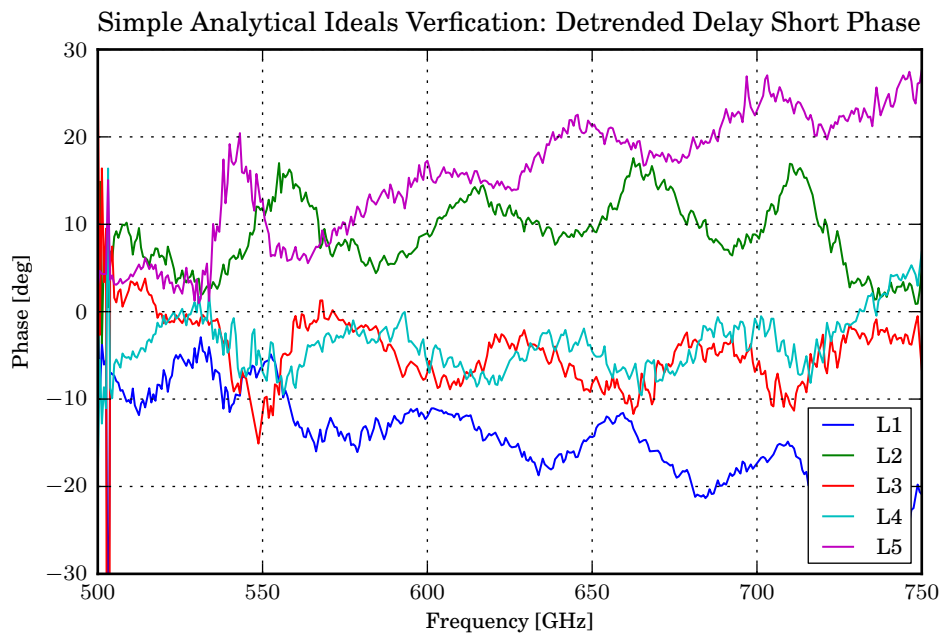
Figure A.20.:
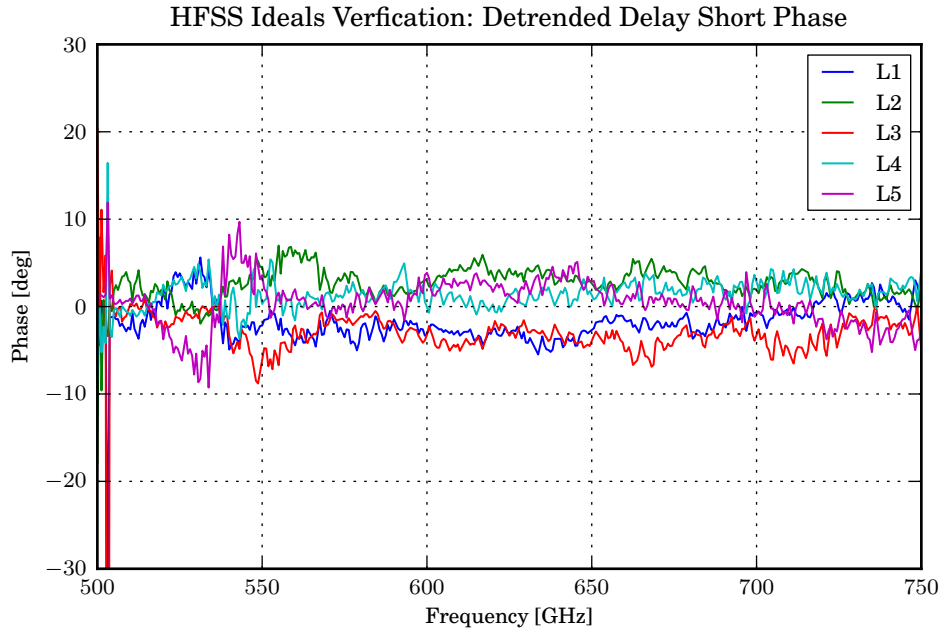
Figure A.21.:

B. The Smith Sphere

B.1. Foundation

B.1.1. Introduction

The Smith Chart

A ubiquitous tool in microwave engineering is the Smith Chart. This chart is simply a plot of rectangular impedance contours on the reflection coefficient plane, as shown in Figure Fig. B.1. Despite the field's widespread adoption of computer aided design (CAD) techniques, the chart is still omnipresent in all design and measurement systems. This begs the question; what is so special about this chart? I attribute the success of the Smith Chart to two characteristics.

- ability to provide the user with a geometrical intuition for otherwise elusive problems

- practical tool for calculations
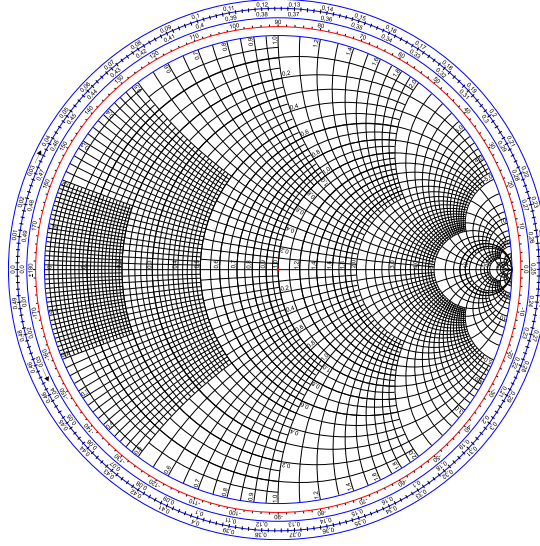
Figure B.1.: A Smith Chart

Since the creation of the smith charts in 1940's it has been recognized that the impedance and reflection coefficient can be projected onto the Riemann sphere[73]. This was also noted in the 'geometric-analytic network theory' put forth by Bolinder[74]. In both cases, however, a full exploration in to the uses of the sphere directly were never made. The major advantage of such a representation is that the entire space of passive and active devices exists within a closed space.

More recently, others have begun to recognize the possibilities of using the Riemann sphere for microwave analysis. However, practical application of higher dimensional topologies such as the Riemann sphere will remain out of reach until the proper algebraic language is used. This language is Geometric Algebra[75].

The Smith Sphere

In this paper, we map the complex impedance and reflection coefficient planes onto a Riemann sphere using stereographic projection and explore some of the consequences. For lack of a better description, we term the this model the *Smith Sphere*.

The model is supported by a three dimensional geometric algebra (GA)[75]. [Say more]

The Smith Sphere has many advantages, both intuitively and computationally, over existing 2D charts. Some benefits of such a construction are

1. a single point can be simultaneously interpreted as an impedance or reflection coefficient

2. the relationship between reflection coefficient and impedance is linearized

3. effects of all common network elements can be represented as circular paths on sphere

4. numerous 2D charts relating network quantities can be generated to suit the problems at hand.

In summary, the Smith Sphere provides a natural space for understanding transmission line phenomenon and provides a flexible model for visualizing circuit behavior. The main draw-back is the algebraic complexity needed for dealing with 3D space. This is unavoidable, as the complexity is what provides the computational power needed to make usage tractable.

The Conformal Smith Surface*

The Smith Sphere is a projective geometry model constructed for a space supporting a one-port network. We shall call this One-Port Projective Space (1PS). While such a projective space provides many benefits to living in flatland, even more convinces are gained by moving to a conformal geometric algebra of one dimension higher. Such a space will be of four dimensions and of mixed signature. The Smith Sphere then becomes a surface existing in 4-dimensions, which is impossible to visualize.

Because of this visualization difficulty, the *Conformal Smith Surface* will be introduced after the Projective Smith Sphere, once some geometrical intuition is formed and the algebra is made familiar. It just so happens that the appropriate conformal GA for a one-port network is the same algebra used for Space-Time, commonly referred to as Space-Time-Algebra (STA)[31]. This is pointed out in[34]. For classification in Geometric Network Theory, this will be called a One-Port Conformal Space (1CS). The Conformal Smith Surface will be worked out in a future paper.

Another problem yet to be dealt with is extension of the Smith Sphere to two-port networks and above. The lack of vector-spinor isomorphism (expressed $\mathbb{G}_2^+ = \mathbb{G}_2^-$) is higher dimensional GA's may force some concepts to be altered, or a different conformal split to be used.

B.1.2. Stereographic Projection

Note to the reader

The mathematics in this paper are expressed in the language of Geometric Algebra (GA). Although GA is essential in making this a usable model, I have tried through the use of graphics and language, to make the concepts in this paper available to the casual reader, especially those not fluent in this subject of GA. In fact, most of this paper could probably be read by a microwave engineer by the pictures alone.

This section presents the mathematical foundations for the rest of the model. But, if the reader is interested to see the *why* before they put forth the effort to understand the *how,* then this section (sec. B.1.2) can be skipped, and returned to later.

General Concept

The mathematical idea behind projective geometry is that a 2 dimensional quantity can be represented by a vector in 3 dimensional space. The purpose of such a representation is that certain relationships and transformations become simpler. Think of working with a globe as opposed to a flat map of the earth. The basic steps are as follows

1. Start with the original 2D vector space.

2. Add an extra dimension, orthogonal to the first two.

3. Using a projective relationship, *map* the 2D plane to a desired surface (in this case a sphere).

The fact that the point in the new 3D space is forced to lie on the prescibed surface eliminates the extra degree of freedom introduced by the additional dimension. From

a mathematical point of view, the Riemann sphere is attractive because the point at infinity becomes an ordinary point on the sphere. Also, the surface is closed.

In this section the impedance plane is mapped onto the sphere, but there is nothing specific in the derivation to the concept of impedance, other than the notation.

Vector Dual of Impedance Spinor

Before we can construct the stereographic projection, we must convert one-port impedance spinor into its vector dual.

As described in *Geometric Network Theory*[1], the electrical phenomena relevant to a one-port network can be represented in the 2 dimensional GA, $\mathbb{G}^2$. The algebra is generated by two orthonormal vectors, $\mathbf{e_r}$ and $\mathbf{e_x}$ with positive signature. The entire algebra is given by,

$$\mathbf{e_r^2} = \mathbf{e_x^2} = 1, \qquad \mathbf{e_r}\mathbf{e_x} = I$$

$$\{\alpha, \quad \mathbf{e_r}, \mathbf{e_x}, \quad I\}$$

Where I is the psuedo-scalar for the port-plane, or Z-plane, of $\mathbb{G}^2$. Within this algebra, a network can be represented by a Impedance Spinor,

$$Z = \alpha + \beta I$$

---

[1]This document is a work in progress

In $\mathbb{G}^2$, the every spinor can be represented by a vector dual ($\mathbb{G}_2^+ = \mathbb{G}_2^-$). This relation is fixed by choosing an reference vector, which is interpreted as the resistance axis, and pre-multiplying. [Figure out if this should be Z^2 or Z]. Thus the vector dual of the Impedance Spinor, denoted $\mathbf{z}$, is given by

$$
\begin{aligned}
\mathbf{z} &= \mathbf{e_r} Z \\
&= \mathbf{e_r} \alpha + \beta \mathbf{e_r} \mathbf{e_r} \mathbf{e_x} \\
&= \alpha \mathbf{e_r} + \beta \mathbf{e_x}
\end{aligned}
$$

$\mathbf{z}$ lies within the Z-plane, expressed by,

$$
\mathbf{z} \wedge I = 0
$$

So we have a two dimensional real vector space which represents a one-port network.

Normalization

To make what follows align with conventional theory I will assume that the impedance vector, $\mathbf{z}$, can be normalized to some real impedance value. It is mathematically convenient to choose the characteristic impedance of the transmission line as the normalizing impedance, which is traditionally a real value. However, for wave-guiding transmission lines this is not the case. The effect of arbitrary normalization (division by a vector) will be dealt with later. [Give better argument].

Projection onto Sphere

Next, the Z-plane is expanded to a three dimensional space by the addition of a vector, labeled $\mathbf{e_s}$, that is orthogonal to the first two and has positive signature. The resultant space is $\mathbb{G}^3$, and its vector basis is illustrated below.
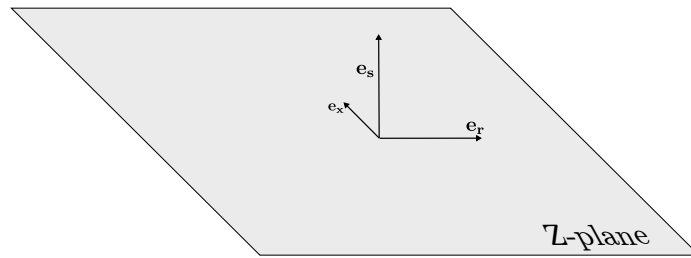
Figure B.2.: Vector Basis Set for the Smith Sphere

To eliminate the extra degree of freedom, the entire Z-plane is mapped to the surface of a unit sphere through stereographic projection, defined as follows;

Let '$\mathbf{z}$' be a point in the Z-plane, then '$\mathbf{p}$' will be the corresponding point lying on the surface of the sphere. The point is determined by drawing a 'ray' connectg the projection point, which is $\mathbf{e_s}$, to a point in the Z-plane, $\mathbf{z}$. The intersection of such a 'ray' with the surface of the sphere yields $\mathbf{p}$. The geometrical relationship is illustrated below. When $|\mathbf{z}| < \mathbf{1}$, the ray is be projected through the Z-plane, onto the interior of the sphere.
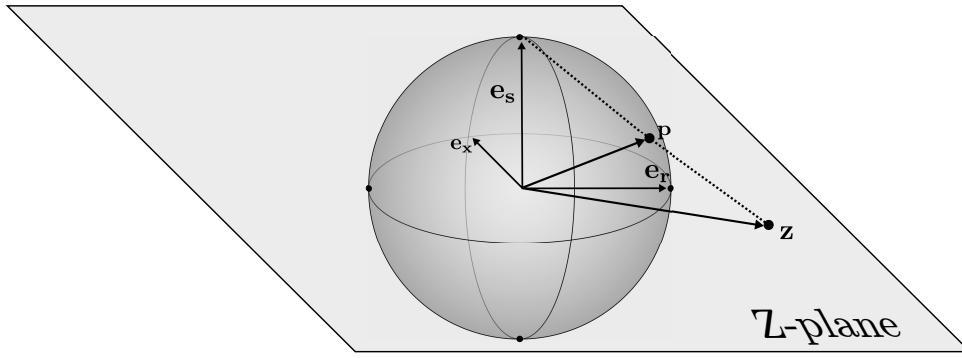
Figure B.3.: Stereographic Projection of Impedance plane onto Sphere

In order to transform operations onto the sphere, an explicit relationship between the points $\mathbf{z}$ and $\mathbf{p}$ must be found[2]. From Figure Fig. B.3 It is geometrically obvious that,

$$\mathbf{p} = \mathbf{e_s} + \lambda\left(\mathbf{z} - \mathbf{e_s}\right) \tag{B.1}$$

where λ is some scalar. The conditions that $\mathbf{p}$ lie on a unit sphere as well as ortho-normalized condition on $\mathbf{e_s}$ provides the following constraints,

$$\mathbf{p^2} = \mathbf{1} \qquad \mathbf{e_s^2} = \mathbf{1} \qquad \mathbf{e_s} \cdot \mathbf{z} = 0$$

enforcing the first condition, sets up the needed constraints as follows,

---

[2]This proof was found in [76], and in a slightly different way in [35].

$$
\begin{aligned}
\mathbf{p} \cdot \mathbf{p} &= \left(\mathbf{e_s} + \lambda\left(\mathbf{z} - \mathbf{e_s}\right)\right) \cdot \left(\mathbf{e_s} + \lambda\left(\mathbf{z} - \mathbf{e_s}\right)\right) \\
1 &= \mathbf{e_s^2} + \lambda^2 \left(\mathbf{z} - \mathbf{e_s}\right) \cdot \left(\mathbf{z} - \mathbf{e_s}\right) + 2\lambda\left(\mathbf{e_s} \cdot \mathbf{z} - \mathbf{e_s^2}\right) \\
1 &= \mathbf{e_s^2} + \lambda^2 \left(\mathbf{z^2} - \mathbf{e_s^2} - \mathbf{2e_s} \cdot \mathbf{z}\right) + 2\lambda\mathbf{e_s} \cdot \mathbf{z} - 2\lambda\mathbf{e_s^2} \\
1 &= 1 + \lambda^2 \left(\mathbf{z^2} - \mathbf{e_s^2}\right) - 2\lambda \\
\lambda &= \frac{2}{\mathbf{z}^2 + 1}
\end{aligned}
$$

putting this into equation B.1 provides the functional relationship for a point on the sphere in terms of a point on the Z-plane.

$$
\boxed{\mathbf{p} = P_z\left(\mathbf{z}\right) = \mathbf{e_s} + \frac{2}{\mathbf{z}^2 + 1}\left(\mathbf{z} - \mathbf{e_s}\right)} \tag{B.2}
$$

Where the 'P_z' is understood as projection function from Z-plane. Although the proof was based on geometry of mapping a plane to a sphere, the result holds true for any dimension of $\mathbf{z}$[35].

Projection onto Plane

It will also be necessary to map a point on the sphere to a point on the plane. This can be solved with the following. [3] The fact that the points $\mathbf{e_s}, \mathbf{p}$, and $\mathbf{z}$ all lie in a

---

[3][Im actually not sure how this works entirely, how am i enforcing $\mathbf{p}$ to lie on a the sphere?]

straight line enforces the following equality of their outer products.

$$\mathbf{e_s} \wedge \mathbf{p} + \mathbf{p} \wedge \mathbf{z} = \mathbf{e_s} \wedge \mathbf{z} \tag{B.3}$$

An illustration is shown Figure Fig. B.4. Note that this equality holds true despite the order of $\mathbf{z}$ and $\mathbf{p}$ along the line. Sketching the specific cases; $|\mathbf{z}| < \mathbf{1}$, $|\mathbf{z}| > \mathbf{1}$ is useful.

From this, $\mathbf{p}$ can be broken up into components parrallell with perpendicular to the vector $\mathbf{e_s}$.

$$\mathbf{p} = \underbrace{\mathbf{e_s}\left(\mathbf{e_s} \cdot \mathbf{p}\right)}_{\mathbf{p}_{\parallel}} + \underbrace{\mathbf{e_s}\left(\mathbf{e_s} \wedge \mathbf{p}\right)}_{\mathbf{p}_{\perp}}$$

Using this in (B.3),

$$\mathbf{e_s} \wedge \left(\mathbf{p}_{\parallel} + \mathbf{p}_{\perp}\right) + \left(\mathbf{p}_{\parallel} + \mathbf{p}_{\perp}\right) \wedge \mathbf{z} = \mathbf{e_s} \wedge \mathbf{z}$$

$$\mathbf{e_s} \wedge \mathbf{p}_{\perp} + \mathbf{p}_{\parallel} \wedge \mathbf{z} = \mathbf{e_s} \wedge \mathbf{z}$$

$$\mathbf{e_s}\mathbf{p}_{\perp} + \mathbf{p}_{\parallel}\mathbf{z} = \mathbf{e_s}\mathbf{z}$$

$$\mathbf{e_s}\mathbf{z} - \mathbf{p}_{\parallel}\mathbf{z} = \mathbf{e_s}\mathbf{p}_{\perp}$$

Multiplying both sides by $\mathbf{e_s}$ and solving for $\mathbf{z}$,

$$\mathbf{z} - \mathbf{e_s p_{\parallel} z} \;=\; \mathbf{p_{\perp}}$$

$$(1 - \mathbf{e_s} \cdot \mathbf{p})\,\mathbf{z} \;=\; -\mathbf{p_{\perp}}$$

Thus,

$$\boxed{\mathbf{z} = P_z^{-1}\left(\mathbf{p}\right) = \left(\frac{1}{1 - \mathbf{e_s} \cdot \mathbf{p}}\right)\mathbf{p_{\perp}}} \tag{B.4}$$
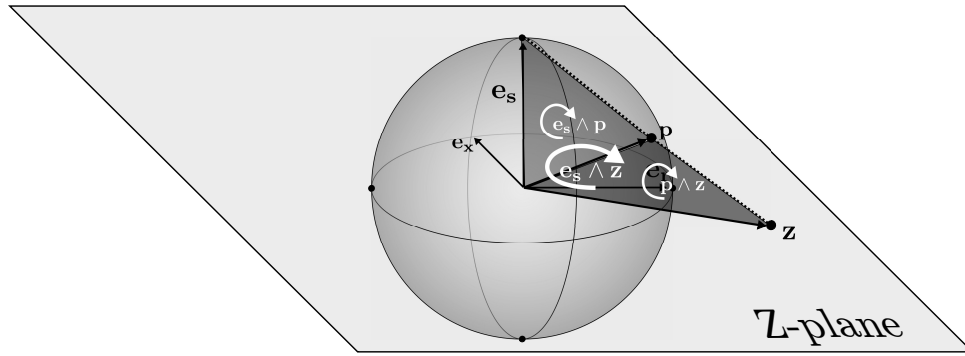

Figure B.4.: Stereographic Projection of Sphere onto Impedance Plane

Formula (B.2) and (B.4) are sufficient to start making some quantitative calculations, but first, a guided tour of the Smith Sphere is given.

B.1.3. The Smith Sphere

An illustration the Smith Sphere, with impedance elements labeled, is shown in Figure Fig. B.6. The labels will be explained in what follows, although someone familiar with the smith chart will probably understand all parts of the figure by staring at it.
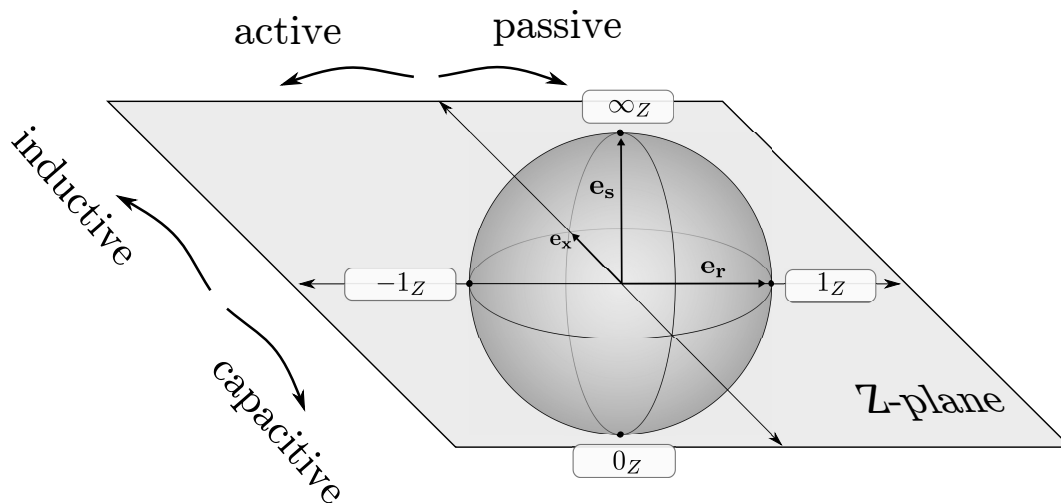
Figure B.5.: The Smith Sphere with Impedance labels

Hemispheres

It is useful to categorize the impedance plane in half-spaces, because this division corresponds to important physical distinctions. Through the projection described above and illustrated in figure Fig. B.6, it can be seen that these half-spaces map to hemispheres on the Sphere.

The right hand side of the Impedance plane corresponds to an impedance with positive resistance, which indicates a passive network. Likewise, the left-hand side corresponds to active networks. These two half-planes map to the east and western hemisphere of the smith sphere.

Assuming the normalizing impedance is real, then the top and bottom half-spaces of the impedance plane represent inductive and capacitive networks, respectively. These map to the back and frontal hemispheres of the Smith Sphere, as seen from the viewpoint shown in Figure Fig. B.6.

Poles

The poles of the Smith Sphere can be labeled for general usage because we are working with normalized impedance values.

- The North pole, which is also the projection pole, represents an infinite impedance, or 'open' circuit.

- The South pole represents a zero impedance, or 'short' circuit.

- The Eastern pole represents a normalized impedance of '1', which is a 'match'.

- The Western pole represents an anti-match, which I dont know how to interpret yet.

Although not explicitly labeled, the back-facing pole is an inductance of normalized value 1, and the front facing pole is a capacitance of normalized value 1.

B.2. Transformations

B.2.1. Impedance and Reflection Coefficient

Determination from fixed points (the poles)

One way to study transformations is through their effect on specific points. Examining the value of the for impedance and reflection coefficient at the poles reveals the relationship between them. The functional relationship is given by

$$\mathbf{s} = \frac{\mathbf{z} - 1}{\mathbf{z} + 1}$$

From this we can evaluate the values for both $\mathbf{z}$ and $\mathbf{s}$ at the poles.

pole	$\mathbf{z}$	$\mathbf{s}$	name
north	∞	1	open
eastern	1	0	match
south	0	-1	short
western	-1	∞	anti-match

Table B.1.: pole table

The relationship between the ordering of the poles for $\mathbf{z}$ and $\mathbf{s}$ indicates that the impedance is related to the reflection coefficient by a 90deg rotation of the sphere, about the $\mathbf{e_x}$ axis. This transform is part of the orthogonal group in $\mathbb{G}^3$, and can be written.

$$R_{zs} = e^{-\frac{\pi}{4}\mathbf{e_r}\mathbf{e_s}} \tag{B.5}$$

This transform can be used in a few ways. If the projection point remains fixed at the north pole, then the smith sphere must rotate. This is achieved by applying the 'zs'-Rotor, given in equation B.5, to a point on the Sphere $\mathbf{p}$. This can be unambiguously expressed, by subscripting the '$\mathbf{p}$'-points on the sphere to indicate their interpretation.

$$\mathbf{p_s} = R_{zs}\mathbf{p_s}R_{zs}^{\dagger}$$

Alternatively, you can think of the Sphere remaining fixed and the projection point rotating towards the western hemisphere. This is achieved by creating an appropriate projection function similar to equation B.2. Written loosely as,

$$
\begin{aligned}
P_s &= R_{zs} P_z R_{zs}^\dagger \\
&= -\mathbf{e_r} + \frac{2}{\mathbf{s}^2 + 1} \left(\mathbf{s} + \mathbf{e_s} \right)
\end{aligned}
$$

From this projection point the reflection coefficient plane, labeled the S-plane, can be drawn as shown in figure Fig. B.6. The advantage of the later concept is that for each point on the sphere, $\mathbf{p}$, there simultaneously exists a value in the S-plane and Z-plane. In other words,

$$
\mathbf{p} = P_s \left(\mathbf{s} \right) = P_z \left(\mathbf{z} \right)
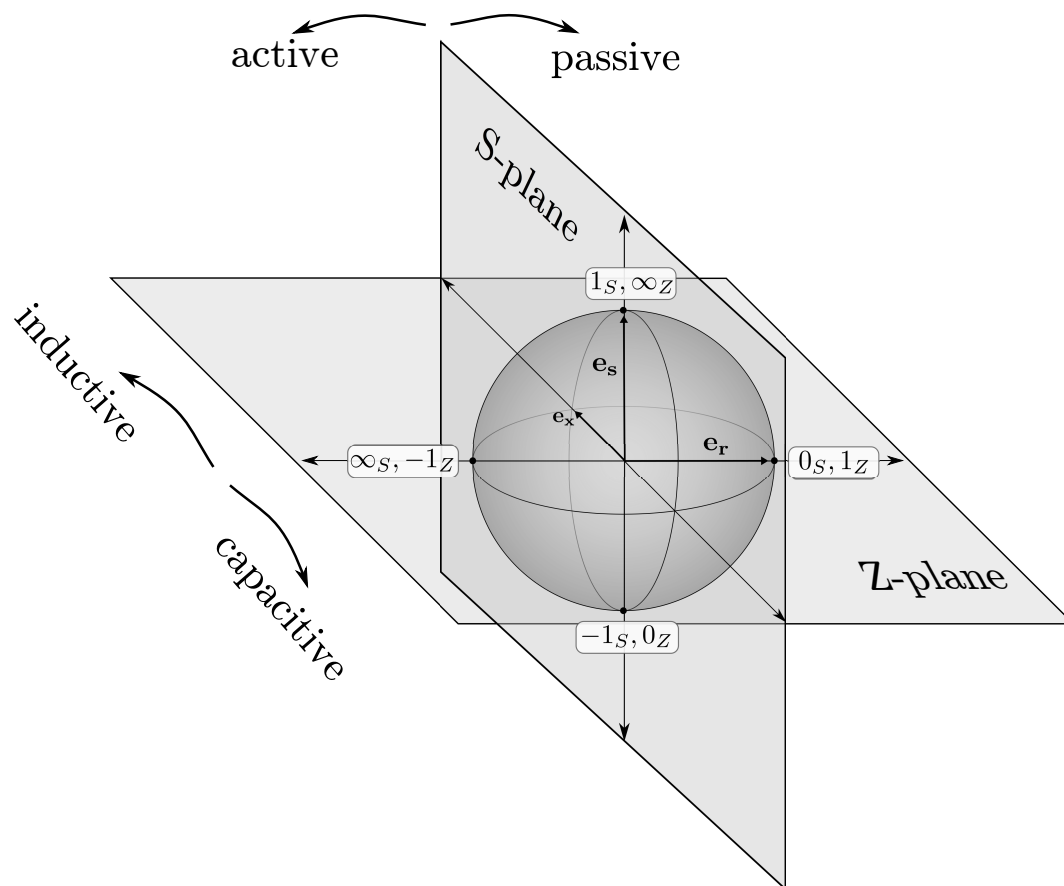$$

Figure B.6.: The Smith Sphere

C. Variational Expressions for Waveguide Junction Admittance

C.0.2. Warning and Disclaimer

This chapter contains useful developments but has **not** been throughly reviewed. Although the variational method discussed has been tested by comparison with experiment, the theory may still contain serious errors. Please do not reference or rely upon it.

C.0.3. Background

The problem of determining the reflection off a generic cylindrical waveguide junction can be solved using the Generalized Scattering Matrix (GSM) (aka Modal Analysis, Eigen-mode Analysis). This approach is described in various forms by Jaimeson and Whinnery, wexler, claricoats and slin, and others. The method makes no assumptions about the field in the junction, and therefore produces a highly accurate solution. The drawback is that in general, there is no closed form solution,

so every junction must be solved individually, and relies on a computer's ability to invert large matrices.

An alternative approach is to use older variational methods, as described by marcuvitz, collins, and harrington. These have the advantage of being reasonably accurate, fast to evaluate, and intuitive. The simplicity of the variational methods results from their precision: they solve only for what information is desired (much like mason's rule in signal flow graph analysis). The drawback is that variational expressions produce approximate solutions. However a bound on the in-acuracy can be determined. and additional discontinuities (not simple terminations) are far away. Both these conditions are met by the case of a terminated waveguide junction.

The exact solution produced by GSM is found by enforcing all possible reactions to be conserved, given by the following,

$$\left\langle \mathbf{E}^{I}, h^{I} \right\rangle \Big|_{z=0} = \left\langle \mathbf{E}^{II}, h^{I} \right\rangle \Big|_{z=0}$$

$$\left\langle e^{II}, \mathbf{H}^{I} \right\rangle \Big|_{z=0} = \left\langle e^{II}, \mathbf{H}^{II} \right\rangle \Big|_{z=0}$$

This formulation produces enough equations to solve for all unknown mode amplitudes, and thus the entire scattering matrix for the junction.

The variational approach yields an approximate solution by assuming and aperture field, and then enforcing the conservation of reaction across the junction between the two waveguides for that field.

$$\left\langle I_{\sim}, I_{\sim} \right\rangle \Big|_{z=0} = \left\langle II_{\sim}, II_{\sim} \right\rangle \Big|_{z=0}$$

As long as the assumed aperture field is reasonable, the calculated value for the admittance will be close to the correct value due of the stationary nature of the reaction. Furthermore, the problem may be formulated in two different ways: the aperture-field form or the obstacle-current form. The two cases provide upper and lower bounds on the value of the input admittance, thus providing a metric of accuracy for the approximate solution obtained.

Furthermore, more accurate solutions can be found by representing the aperture feild itself as a weighted sum of basis functions, and exploited the stationary nature of the solution to solve determine the values of the weights. This is refered to as the Raliegh-Ritz method, which frequenty coencides with the Galerkin method, when the basis set is the eigen space of the aperture.

C.1. Aperture-Field Formulation

C.1.1. Intro

Imagine two cylindrical waveguides which are butt together, forming a junction as shown in figure (Fig. C.1) . We want to determine the reflection coefficient off such a junction looking for the dominant mode of one of the guides. Due to the parallel transmission-line nature of waveguides, the easiest way to formulate the effects of such discontinuity is through admittances.

The reflection coefficient for the single mode p as seen looking in from region I is given in terms of the normalized input admittance for mode p by,

$$\Gamma_p = \frac{1 - Y_{in,p}^{I'}}{1 + Y_{in,p}^{I'}}$$

Therefore,

We seek an expression for the normalized input admittance for a single mode looking in from one side of the junction.
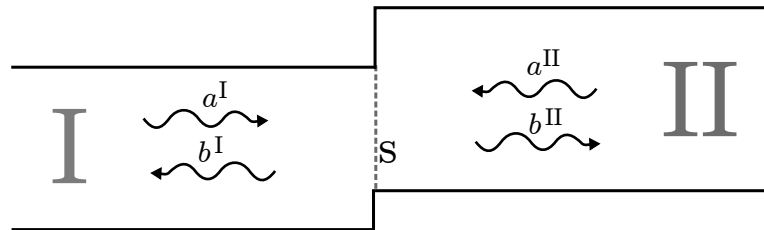

Figure C.1.: Illustration of waveguide junction.

The basic physical law on which our analysis is based upon is the conservation of reaction across the junction. Frequently this is written with the inner-product notation used by Rumsey and Harrington which is,

$$\langle I, I \rangle = \langle II, II \rangle$$

Take the $'z'$-axis to be the longitudinal axis along which energy propagates, and transverse to the junction interface. The reaction *accross* the interface, evalutated in the aperture at $z = 0$ is then

$$\iint \mathbf{E} \cdot \mathbf{J} \stackrel{I}{-} \mathbf{H} \cdot \mathbf{M} \bigg|_{z=0} = \iint \mathbf{E} \cdot \mathbf{J} \stackrel{II}{-} \mathbf{H} \cdot \mathbf{M} \bigg|_{z=0} \tag{C.1}$$

The variational technique, relies on the author to guess at the aperture field. The assumed aperture field is then decomposed into a sum of weighted wave-modes in both regions, and the conservation of reaction is then enforced.

1. Assume a field in the aperture, this is the steady-state system behavior at the junction.

2. The conservation of power (or reaction) accross the aperture, combined with the boundary conditions of each region, provides enough constraints to calculate how much power must exist in each mode, in each region.

3. tbd.

There are two ways to forumulate a variational formula for normalized input admittance; the aperture-field formulation, and the obstacle-current formulation. The

difference between the two is that 'the former assumes a knowledge of tangential electric field, and the latter assumes knowledge of the discontinuity of magnetic field.'[1] As stated earlier, one will provide a lower bound and the other an upper bound on the value for admittance.

The combination of both of these formulations allong with the crucial fact that in a given region (read set of boundary conditions) equalities can be enforced on a per-mode basis (due to linearity of the system), may lead to exact results. I believe this is how Eisenhart and Khan's method works, but I havent explicitly convinced myself.

C.1.1.1. Outline

An outline of the analysis is as follows.

ust be made from a sum of modes which are subject to the boundary conditions of each regions.

1. Represent the transverse fields as a sum of weighted basis functions (functional decomposition).

$$
\mathbf{E_t} = \sum_i c_i e_i \qquad c_i = \iint \mathbf{E_t} \cdot e_i
$$
$$
\mathbf{H_t} = \sum_i c_i h_i \qquad c_i = \iint \mathbf{H_t} \cdot h_i \tag{C.2}
$$

2. Assume a reasonable guess for the aperture field, and enforce conservation of reaction in the aperture.

$$
\iint \mathbf{E} \overset{I}{\times} \mathbf{H} \bigg|_{z=0} = \iint \mathbf{E} \overset{II}{\times} \mathbf{H} \bigg|_{z=0}
$$

---

[1]Roger F. Harington - Time Harmonic Electromagnetic Fields: section 4-10

3. Using per-mode relationship between fields and admittance, serperate and solve for normalzied input admittance for the mode of interest.

C.1.2. Analysis

C.1.2.1. Waveguide Modes as Parrallel Transmission Lines

$$
\begin{aligned}
\mathbf{E} &= \overrightarrow{E}\,(x,y,z) \\
&= \hat{x} \cdot E_x\,(x,y,z) + \hat{y} \cdot E_y\,(x,y,z) + \hat{z} \cdot E_z\,(x,y,z) \\
&= \hat{x} \cdot E_x\,(x)\,E_x\,(y)\,E_x\,(z) +
\end{aligned}
$$

$$
\mathbf{E}\,(x,y,z) = \underbrace{\mathbf{E_t}\,(x,y,z)}_{\text{transverse}} + \underbrace{E_z\,(x,y,z)}_{\text{longitudinal}}
$$

$$
\begin{aligned}
E_t\,(x,y,z) &= V(z) \cdot e(x,y) \\
H_t\,(x,y,z) &= I(z) \cdot h(x,y)
\end{aligned}
\tag{C.3}
$$

The motivation to seperate the transverse and longitudinal functional dependence is that the two boundary conditions are characteristically different. The boundary conditions in the transverse direction lead to a discrete set of transverse eigen-values and the to-be-determined boundary conditions in the longitudinal conditions, are enforced later through the input admittance terms.

C.1.2.2. Functional decomposition

There are two main mathematical models for the study of wave propagation; the fourier optics model and the Huygens-Fresnel principle. The interpretations produced by each, and their relationship is described by an unkown author on wikipedia,

> Fourier optics is the study of classical optics using Fourier transforms and can be seen as the dual of the Huygens-Fresnel principle. In the latter case, the wave is regarded as a superposition of expanding spherical waves which radiate outward from actual (physically identifiable) current sources via a Green's function relationship. In Fourier optics, by contrast, the wave is regarded as a superposition of plane waves which are not related to any identifiable sources; instead they are the natural modes of the propagation medium itself. [2]

In this paper we will use the fouier optics approach, usually refered to as modal analysis when applied to guided wave. An arbitrary transverse field can be equivalently represented as a sum of modes. The general name for such a technique is functional decomposition.

$$
\begin{aligned}
\mathbf{E_t} &= \sum_i c_i e_i \qquad c_i = \iint \mathbf{E_t} \cdot e_i \\
\mathbf{H_t} &= \sum_i c_i h_i \qquad c_i = \iint \mathbf{H_t} \cdot h_i
\end{aligned}
\tag{C.4}
$$

Where c_i is the constant weighting coefficient for the ith mode. Mathematically, the c_i's are the amplitudes of the *transverse mode functions* e_i, h_i, and the transverse mode functions are the eigen-functions of Maxwell's wave equation subject to the

---

[2]http://en.wikipedia.org/wiki/Fourier_optics

boundary conditions of the waveguides.

This 'sum of modes' representation is much like the way an apritrary band-limited time signal can be fully represented by a Continuous Time Fourier Transform.

The electric and magnetic eigen-functions are related by,

$$e_i = h_i \times \hat{\mathbf{z}} \tag{C.5}$$

C.1.2.3. Mode-Voltages and Currants

In general, the total field in either region is a summation of forward and reflected traveling waves.

$$\mathbf{E_t} = \sum_i \left(a_i e^{j\gamma z} + b_i e^{-j\gamma z} \right) e_i$$

$$\mathbf{H_t} = \sum_i \left(Y_i a_i e^{j\gamma z} - Y_i b_i e^{-j\gamma z} \right) h_i$$

Where a_i's are the electric field mode-amplitudes for *forward* traveling waves, b_i's are the electric field mode amplitudes for *reverse* traveling waves, and Y_i's are the *characteristic* admittance for mode $'i'$. At the junction, $z = 0$, the above expression simplifies to just the transverse field components,

$$\mathbf{E}_t = \sum_i (a_i + b_i)\, e_i \tag{C.6}$$

$$\mathbf{H_t} = \sum_i Y_i\, (a_i - b_i)\, h_i$$

Using Kurakowa's definition of scattering waves, provides a useful relation between the *traveling wave* mode-amplitudes and *total wave* mode-amplitudes.

$$a_i = \frac{1}{2} k_i\, (V_i + Z_i I_i)$$

$$b_j = \frac{1}{2} k_j\, (V_j - Z_j I_j)$$

then equation C.6 becomes,

$$(a_i + b_i) = k_i\, V_i$$

$$Y_i\, (a_i - b_i) = k_i\, I_i$$

V_i is the the ith *mode-voltage,* and I_i is the ith *mode-current.* These quantities are proportional to the total fields for a given mode. Comparing the above with functional decomposition given in equation C.4, and setting the scaling factor of $k_i = 1$,

$$\mathbf{E_t} = \sum_i \underbrace{(a_i + b_i)}_{V_i} e_i \tag{C.7}$$

$$\mathbf{H_t} = \sum_i \underbrace{Y_i (a_i - b_i)}_{I_i} h_i$$

$$V_i = \iint \mathbf{E_t} \cdot e_i$$

$$I_i = \iint \mathbf{H_t} \cdot h_i$$

The relation of total fields and scattering parameters will connect reflection coefficient with aperture admittance.

C.1.2.4. Enforcing Conservation of Reaction

The self-reaction of a field with itself just to one side of the junction is,

$$\langle I, I \rangle = \iint \mathbf{E} \cdot \mathbf{J} \overset{I}{-} \mathbf{H} \cdot \mathbf{M}\Big|_{z=0}$$

$$= - \iint \mathbf{E} \overset{I}{\times} \mathbf{H^*}\Big|_{z=0}$$

Substituting in the waveguide mode basis function expansion,

$$
\begin{aligned}
\iint_{ap} \mathbf{E} \times \mathbf{H} ds &= \iint_{ap} \left\{ \sum_i V_i e_i \times \sum_j I_j h_j \right\} \\
&= \sum_i \sum_j V_i I_j \iint_{ap} e_i \times h_j \\
&= \sum_j \sum_i V_i I_j \iint_{ap} e_i \cdot e_j ds
\end{aligned}
$$

Where the relation given by equation C.5 has been used in the last step. The integral is evaluated over the cross-section of the region of interest at $z = 0$. If the waveguide modes where carefully crafted such that they form an orthonormal set, the last expression simplifies to a single sum.

$$
\begin{aligned}
&= \sum_j \sum_i V_i I_j \underbrace{\iint_{ap} e_i \cdot e_j ds}_{\delta_{ij}} \\
&= \sum_j \sum_i V_i I_j \delta_{ij} \\
&= \sum_i V_i I_i
\end{aligned}
$$

Enforcing the conservation of reaction over the aperture, and using the ortho-normal basis set property given above,

178

$$\langle I, I \rangle = \langle II, II \rangle$$

$$\iint \mathbf{E} \overset{I}{\times} \mathbf{H} \Big|_{z=0} = \iint \mathbf{E} \overset{II}{\times} \mathbf{H} \Big|_{z=0} \tag{C.8}$$

$$\sum_i \overset{I}{V_i} I_i = \sum_j \overset{II}{V_j} I_j \tag{C.9}$$

The mode voltage and currents for the assumed field at either side of the junction are found by,

$$V_i = \iint \mathbf{E_{ap}} \cdot e_i$$

$$I_i = \iint \mathbf{H_{ap}} \cdot h_i \tag{C.10}$$

The boundary conditions are enforced both by the bounds of the integral, and the forms of the basis function.

C.1.2.5. Reformulating in Terms of Input Admittance

The relation between the *normalized input admittance* for mode i (*normalized* conveyed by the apostrophe) and the mode-voltage and mode-currents is found through application of transmissionline theory.

$$Y'_{in,i} = \frac{1 - S_{ii}}{1 + S_{ii}} = \frac{1 - \frac{b_i}{a_i}}{1 + \frac{b_i}{a_i}} = \frac{(a_i - b_i)}{(a_i + b_i)} = \frac{1}{Y_i} \frac{I_i}{V_i} \tag{C.11}$$

Looking back at equation C.9, it can be seen that there is a choice now to substitute for I_i or V_i. This results of this choice is that one will yield and expression in terms of Electric field (known as aperture-field formulation), and the other in terms of the Magnetic field (known as obstacle-current).

Initially we will use the aperture-field formulation. Re-writing equation C.11,

$$I_i = V_i Y_i Y'_{in,i}$$

Substituting this in for the mode-currents in equation (C.9), and seperating the term containing mode of interested, mode 'p' of region I.

$$\sum_i \overset{I}{V_i I_i} = \sum_j \overset{II}{V_j I_j}$$

$$\overset{I}{V_p^2 Y_p Y'_{in,p}} + \sum_{i \neq p} \overset{I}{V_i^2 Y_i Y'_{in,i}} = \sum_j \overset{II}{V_j^2 Y_j Y'_{in,j}}$$

Solving for the normalized input admittance for mode 'p', as seen looking in from region I,

$$\overset{I}{Y'_{in,p}} = \frac{\sum_j \overset{II}{V_j^2 Y_j Y'_{in,j}} - \sum_{i \neq p} \overset{I}{V_i^2 Y_i Y'_{in,i}}}{\overset{I}{V_p^2 Y_p}} \tag{C.12}$$

This expression is in terms of mode-voltage, which is a quantity related to the electric field, as given in equation C.10. Thus, this approach for junction admitance

is commonly reffered to as the 'aperture-field' formulation. Written out explicitly in terms of the aperture electric field,

$$Y_{in,p}^{'I} = \frac{\sum_j \left(\iint \mathbf{E_{ap}} \cdot e_j \right)^2 \overset{II}{Y_j Y_{in,j}'} - \sum_{i \neq p} \left(\iint \mathbf{E_{ap}} \cdot e_i \right)^2 \overset{I}{Y_i Y_{in,i}'}}{\left(\iint \mathbf{E_{ap}} \cdot e_p \right)^2 \overset{I}{Y_p}}$$

The normalized input admittances in the above equations are referenced looking *into* each region. Thus, the above expression makes it is possible to include the effects of a *simple* terminations attached to either waveguide, through these normalized input admittance terms. The term *simple* meaning a termination which can be accurately captured in a per-mode fashion. Examples would be a short, open, or simple load.

C.1.3. Equivalent Circuit

It is worthwhile at this point to pause for a second and draw an equiavlent circuit for for out expression. From equation C.12, one possible equivalent circuit for the is illustated in Figure Fig. C.2 . Its a sum of admittances, one for modes region *I* and one for modes in region *II*. The coupling to each mode, represented by a transformer, is determined by the values of the to-be-determined total mode voltages in each region. The admittances are the un-normalized input impedance looking into each guide, for each mode. Depending on the discontinuity, diferent modes will be excited in different amounts, and this will control whether the junction apears capacitive, inductive, resistive, or lossless.
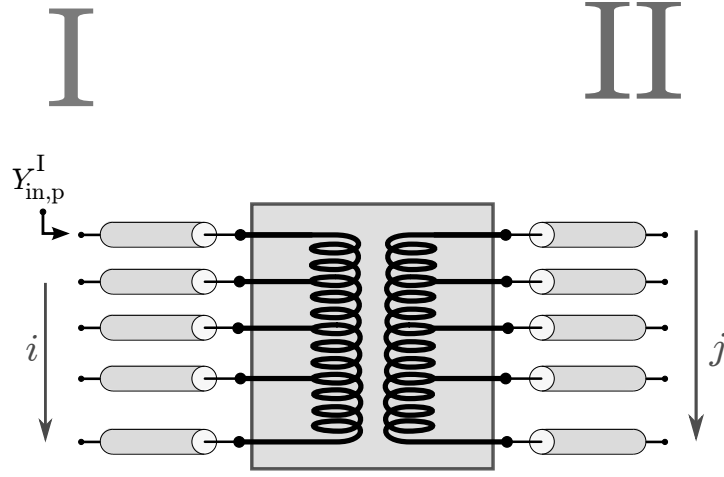
Figure C.2.: Equivalent circuit for waveguide junction.

C.1.4. Special [reasonable] Cases

C.1.4.1. Matched Input

Think practically. We want the reflection off the junction for a single mode. If the junction is excited from region I with a single mode, then it is usually the case that waveguide represented by region I is matched. Meaning there is only one incident mode from region I, our excitation,

$$
a_i^I = \begin{cases} a_p^I & i = p \\ 0 & i \neq p \end{cases}
$$

This condition will effect the sign of the second sum in the numerator of equation, (C.12).

$$Y_{in,i\neq p}^{'I} = \frac{(a_i^I - b_i)}{(a_i + b_i)}\bigg|_{i\neq p} = -1$$

Thus the normalized input admittance for a mode $'p'$ given the input waveguide is matched,

$$Y_{in,p}^{'I} = \frac{\sum_j V_j^2 Y_j^{II} Y_{in,j}^{'} + \sum_{i\neq p} V_i^2 Y_i^I}{V_p^2 Y_p^I} \tag{C.13}$$

C.1.4.2. Matched Input and Output

Having a matched output imposes the condition,

$$a_j^{II} = 0$$

Recall figure Fig. C.1, and note the direction of the wave amplitudes. Due to fact that the current must be oriented in the same direction in both regions.

$$\begin{aligned} I_i^I &= (a_i^I - b_i) \\ I_j^{II} &= (b_j^{II} - a_j) \end{aligned}$$

For matched conditions in region II,

$$Y_{in,j}^{II'} = \frac{(b_j \overset{II}{-} a_j)}{(b_j + a_j)} = 1 \qquad \text{given } a_j^{II} = 0$$

and thus for matched input and ouput,

$$Y_{in,p}^{I'} = \frac{\sum_j V_j^2 Y_j^{II} + \sum_{i \neq p} V_i^2 Y_i^{I}}{V_p^2 Y_p^{I}} \tag{C.14}$$

C.1.5. Details of Evaluating the Field

The original expression for the mode amplitudes given by equation (C.6).

$$\mathbf{E} = \sum_i (a_i + b_i) \, e_i = \sum_i (V_i) \, e_i$$

The quantity V_i is given by,

$$V_i = \iint \mathbf{E} \cdot e_i \, ds$$

All at once this is,

$$\mathbf{E} = \sum_i \left(\iint \mathbf{E} \cdot e_i \, ds \right) \cdot e_i$$

This is an integral transform of the field $\mathbf{E}$ into eigen-space. The bounds on the integral are given by common cross-section of the two guides.

Actually calculating the eigen-space transform is lengthy due to the vectorial nature of the wave equation in 3D space. $\mathbf{E}$ must be decomposed into all permutations of mode-types and transverse eigen-values. Also, unless the summation is a special series which converges to a closed form expression, the number of modes taken into account must be truncated,

$$\mathbf{E} = \sum_r^{tm,te} \sum_m^\infty \sum_n^\infty \left(V_{r,m,n}\right) e \approx \sum_r^{tm,te} \sum_m^M \sum_n^M \left(V_{r,m,n}\right) e$$

Where the total mode voltage is given by

$$V_{r,m,n} = \iint \mathbf{E} \cdot e_{r,m,n} \, ds$$

The eigen-functions $e_{r,m,n}$ are in general different in each region. Known expressions for normalized eigen functions are given by marcuvitz for a variety of common wave-guiding geometries.

C.2. Obstacle-Current Formulation

The obstacle-current formulation is the dual of the aperture-field formulation. Because of this the two analysis are practically identical, and so will be presented consisely without detailed explanaition. The difference in the result, is that the value for junction admittance will be a negative definite.

C.2.1. Analysis

The analysis is formulated by assuming a surface current on the obstacle face instead of the fields in the aperture. By Maxwells equations this surface current must supported by a discontinuity in magnetic field, evaluted over the obstacle. Starting from the conservation of reaction,

$$\langle I, I \rangle = \langle II, II \rangle$$

$$\iint \mathbf{E} \cdot \mathbf{J} \overset{I}{-} \mathbf{H} \cdot \mathbf{M} \bigg|_{z=0} = \iint \mathbf{E} \cdot \mathbf{J} \overset{II}{-} \mathbf{H} \cdot \mathbf{M} \bigg|_{z=0}$$

Evaluating this accross the obstacle face, where the discontinuity of the tangential magnetic field must be supported by a surface current on the obstacle,

$$\mathbf{J}_{\mathbf{obs}}^{I} = \hat{\mathbf{z}} \times \left(\mathbf{H}^{I} - \mathbf{H}^{0} \right)$$

$$\mathbf{J}_{\mathbf{obs}}^{II} = \hat{\mathbf{z}} \times \left(\mathbf{H}^{II} - \mathbf{H}^{0} \right)$$

Where the superscipt 0, refers to the plane seperating the two regions. For an obstacle which is a perfect conductor, the magnetic field in the obstacle face will be zero, $(\mathbf{H}^{0} = 0)$, in which case,

$$\mathbf{J}_{\mathbf{obs}}^{I} = \hat{\mathbf{z}} \times \mathbf{H}^{I}$$

$$\mathbf{J}_{\mathbf{obs}}^{II} = \hat{\mathbf{z}} \times \mathbf{H}^{II}$$

186

Assuming $\mathbf{M} = \mathbf{0}$, and substituting into the above,

$$\iint_{obs} \mathbf{E}^I \cdot \hat{\mathbf{z}} \times \mathbf{H}^I \;\; = \;\; \iint_{obs} \mathbf{E}^{II} \cdot \hat{\mathbf{z}} \times \mathbf{H}^I$$

From this expression it is clear that this formulation can only be used when there is a obstacle in both regions. Because the formulation relies the magnetic field, we will solve for the impedance in terms of mode-current instead of mode-voltage. The mode voltage in terms of mode current is given by equation C.11.

$$V_i = Z'_{in,i} Z_i I_i$$

Starting from equation C.9, substituting the above for the mode-voltage, and solving for normalized input impedance for mode $'p'$ of region I.

$$\sum_i \overset{I}{V_i} I_i \;\; = \;\; \sum_j \overset{II}{V_i} I_i$$

$$I_p^2 Z_p \overset{I}{Z'_{in,p}} + \sum_{i \neq p} I_i^2 Z_i \overset{I}{Z'_{in,i}} \;\; = \;\; \sum_j I_j^2 Z_j \overset{II}{Z'_{in,j}}$$

$$\overset{I}{Z'_{in,p}} \;\; = \;\; \frac{\sum_j I_j^2 Z_j \overset{II}{Z'_{in,j}} - \sum_{i \neq p} I_i^2 Z_i \overset{I}{Z'_{in,i}}}{I_p^2 Z_p}$$

A key difference for the obstacle-current formulation is that the boundary conditions are enforced over different regions than the aperture-field formulation. The

aperture-field formulation enforces continuity of Electric field in the aperture and zero tangential electric field at perfect conductors. These are enforced by placing correct bounds mode-voltage integrals.

The obstacle-current formulation, on the other hand, is in terms of mode-currents which are proportional to the total tangential magnetic field. As stated earlier,

'The [aperture-field form] assumes a knowledge of tangential electric field, and the [obstacle-current form] assumes knowledge of the discontinuity of magnetic field.'-R.F.H.

The total magnetic field is not neccesarily zero at the surface of electrical conductors. Therefore in the obstacle-current case the author assumes a current, which exists on the obstructing face of the junction to support a reasonable magnetic field in each guide.

In short,

- Mode-voltage integrals evaluated in aperture-field formulations are taken over the common cross-section between the two guides.

- Mode-current integrals in obstacle-current formulation are taken over the uncommon cross section in each guide.

C.2.2. Special [reasonable] Cases

C.2.2.1. Matched Input

$$Z_{in,p}^{'I} = \frac{\sum_j I_j^2 Z_j \overset{II}{Z'_{in,j}} + \sum_{i \neq p} I_i^2 \overset{I}{Z_i}}{I_p^2 \overset{I}{Z_p}}$$

C.2.2.2. Matched Output

For Matched input and matched output this becomes

$$Z_{in,p}^{I} = \frac{\sum_j I_j^{2}{}^{II}Z_j + \sum_{i \neq p} I_i^{2}{}^{I}Z_i}{I_p^2{}^{I}Z_p}$$

C.3. Junction Classification

It is advantagous to classify waveguide junctions into two types: *Intersection*, and *Subset*. An illustration of the cross-section of junction for each type is shown below. The reason to make this distinction is because different junction types lead to different mathematical simplifications and thus different equivalent circuits. It is also convinient to classify the bounds on the mode voltage and mode-current integrals for both types of formulations. This is given below in table Tab. C.1.
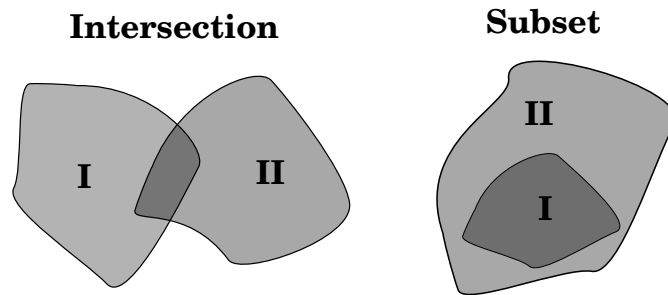
Figure C.3.: Illustration of the two Junction Types

	Aperture-Field		Obstacle-Current	
	Region I	Region II	Region I	Region II
Intersection	$I \bigcup II$	$I \bigcup II$	$I \bigcap (I \bigcup II)$	$II \bigcap (I \bigcup II)$
Subset	I	I	na	$II \bigcap I$

Table C.1.: Integral Bounds Table

C.3.1. Subset Junction

The *Subset* type of waveguide junction is characterized by the aperture cross-section being identical as one of the waveguide's cross-section. For an aperture-field formulation, it may be reasonable to assume that the aperture field is identical to the dominant mode of the exciting waveguide. This simplifies the aperture addmitance expression considerably.

For example, assume the aperture field has the same functional form as mode 'p' of region I. If the modes form an ortho-normal set, the mode voltage for all other modes is zero

$$V_i = \iint E_{ap} \cdot e_i \, ds = \iint e_p \cdot e_i \, ds = \delta_p$$

and thus the second sum in the equation C.13 vanishes,

$$\sum_{i \neq p}^{I} (V_i)^2 Y_i Y'_{in,i} = 0$$

and the expression becomes,

$$Y_{in,p}^{I'} = \frac{\sum_j (V_j)^2 \overset{II}{Y_j} Y_{in,j}'}{(V_p)^2 \overset{I}{Y_p}}$$

This is the method used in Harrington throughout chapter 4.

C.3.2. Intersection Junction

The general case of the intersection junction is handled using the full expression, as given by equation C.13.

C.3.3. Intersection Special Case: Symetric Waveguide Junctions

If the two waveguides have identical cross-sections, then the eigenfunctions and characteristic admittances of each region will be the same.

$$e_i^I = e_i^{II}$$
$$Y_i^I = Y_i^{II}$$

Furthermore, if the junction is symetrical, the common cross-section apears the same looking in from each waveguide, then the mode-voltages squared for each region are the same,

$$\left(V_i^I\right)^2 = \left(V_i^{II}\right)^2$$

and the following simplifications can be made.

$$
\begin{aligned}
Y_{in,p}^{'I} &= \frac{\sum_j V_j^2 Y_j \overset{II}{Y_{in,j}'} - \sum_{i\neq p} V_i^2 Y_i \overset{I}{Y_{in,i}'}}{V_p^2 \overset{I}{Y_p}} \\[2mm]
&= \frac{V_p^2 Y_p \overset{II}{Y_{in,p}'}}{V_p^2 \overset{I}{Y_p}} + \frac{\sum_{j\neq p} V_j^2 Y_j \overset{II}{Y_{in,j}'} - \sum_{i\neq p} V_i^2 Y_i \overset{I}{Y_{in,i}'}}{V_p^2 \overset{I}{Y_p}} \\[2mm]
&= \overset{II}{Y_{in,p}'} + \frac{\sum_{j\neq p} V_j^2 Y_j \overset{II}{Y_{in,j}'} - \sum_{i\neq p} V_i^2 Y_i \overset{I}{Y_{in,i}'}}{V_p^2 Y_p}
\end{aligned}
$$

If both waveguides are matched, the expression simplifies even further,

$$Y_{in,j}^{II} = 1$$

$$Y_{in,j}^{'I} = -1$$

$$
\begin{aligned}
Y_{in,p}^{I'} &= 1 + \frac{\sum_{j\neq p} V_j^2 \overset{II}{Y_j} + \sum_{i\neq p} V_i^2 \overset{I}{Y_i}}{V_p^2 Y_p} \\[2mm]
&= 1 + 2\frac{\sum_{j\neq p} V_j^2 Y_j}{V_p^2 Y_p}
\end{aligned}
$$

C.4. Systematically Improving Assumed Field: Raleigh-Ritz, Galerkin

The raleigh-ritz method is a systematic way of improving the aperture field, and may lead to exact results in some cases. The idea is the represent the aperture field as a set of basis functions, much like the fields in either region. The weights of each of the basis functions are then deteremined by exploiting the stationary nature of the admittance expression.

C.4.1. Special Case: symetric junction

Because of the symplicity of the aperture admittance for symetric junctions it is a good place to illustrate the usage of teh Raleigh-Ritz method. As shown above, the symetric waveguide junction has an aperture admittance of,

$$Y_{in,p}^{I'} = 1 + 2 \cdot \frac{\sum_{j \neq p} V_j^2 Y_j}{V_p^2 Y_p}$$

Explicitly writing out the evaltuation for the mode-voltages, so that we can introduce the basis set expansion,

$$Y_{in,p}^{I'} = 1 + 2 \cdot \frac{\sum_{j \neq p} \left(\iint \mathbf{E^{ap}} \cdot e_j \, ds \right)^2 Y_j}{\left(\iint \mathbf{E^{ap}} \cdot e_p \, ds \right)^2 Y_p}$$

Expressing the aperture field as a sum of weighted basis functions, where the weights are yet to be determined,

$$\mathbf{E^{ap}} = \sum_k c_k e_k^{ap}$$

Substituting this into the previous expression,

$$Y_{in,p}^{I'} = 1 + 2 \cdot \frac{\sum_{j \neq p} \left(\sum_k c_k \iint e_k^{ap} \cdot e_j \, ds \right)^2 Y_j}{\left(\sum_k c_k \iint e_k^{ap} \cdot e_p \, ds \right)^2 Y_p}$$

The stationary nature of the expression is used to determine the weighting coefficients. Specfiically, the admittance is a maximum or minimum around the true value, for small variations in the aperture field function. Given that we are considering $'k'$ terms in the aperture field expansion, the stationary constraints leads to $'k'$ equations, each expressed by

$$\frac{\partial}{\partial c_l} Y_{in,p}^{I'} = 0 \qquad l = 0, 1, 2, ...k$$

$$\frac{\partial}{\partial c_l} \left[1 + 2 \cdot \frac{\sum_{j \neq p} \left(\sum_k c_k \iint e_k^{ap} \cdot e_j \, ds \right)^2 Y_j}{\left(\sum_k c_k \iint e_k^{ap} \cdot e_p \, ds \right)^2 Y_p} \right] = 0 \qquad l = 0, 1, 2, ...k$$

Re-introducing the notation of mode-voltage, with an additional indexing variable for the aperture field basis index, makes this more readable. There is a mode-voltage associated with each combination of aperture field basis functions, and the region's

basis functions.

$$V_{k,,j} = \iint e_k^{ap} \cdot e_j \, ds$$

And the is then written,

$$0 = \frac{\partial}{\partial c_l} \left[1 + 2 \cdot \frac{\sum_{j \neq p} \left(\sum_k c_k V_{k,,j} \right)^2 Y_j}{\left(\sum_k c_k V_{k,,p} \right)^2 Y_p} \right] \qquad l = 0, 1, 2, ... k$$

This quantity is zero, when the numerator vanishes,

$$
\begin{aligned}
0 &= \frac{\partial}{\partial c_l} \left[\sum_{j \neq p} \left(\sum_k c_k V_{k,,j} \right)^2 Y_j \right] \\
&= \sum_{j \neq p} 2 \left(\sum_k c_k V_{k,,j} \right) \cdot V_{l,j} Y_j \\
&= 2 \sum_{j \neq p} \sum_k c_k V_{l,j} V_{k,,j} Y_j \\
&= \sum_k c_k \sum_{j \neq p} V_{l,j} V_{k,,j} Y_j \qquad \text{for all } l \text{ in } k
\end{aligned}
$$

These constraints can be represented as a homogeneous system of equations, containing a known $k \times k$ matrix, and unknown vector $\mathbf{c}$.

$$
\begin{bmatrix}
\sum_{j \neq p} V_{0,j} V_{0,,j} Y_j & \cdots & \sum_{j \neq p} V_{0,j} V_{k,,j} Y_j \\
\vdots & \ddots & \\
\sum_{j \neq p} V_{l,j} V_{0,,j} Y_j & & \sum_{j \neq p} V_{l,j} V_{k,,j} Y_j
\end{bmatrix}
\cdot
\begin{bmatrix}
c_0 \\
c_1 \\
\vdots \\
c_k
\end{bmatrix}
= 0
$$

Such a system has 2 possible solutions; the trivial solution of $\mathbf{c} = 0$, and the infinite set of solutions. Obviously we are interested in the infinite set of solutions, which is called the *null* space of the matrix. The reason it is an infinite set is that the magnitude of the c-vector is not important, only the direction. Therefore there are infinite many $\mathbf{c}$'s which can solve the constraints, but all are scalar multples of each other, and will yield the same results in our expression because it is insensitive to aperture field amplitude.

To make this matrix more concise we can introduce the idea of *sum reaction*. Which is interpreted as the sum of the reactions between a single element of the aperture field basis set, and the entire set of a region's basis set.

$$
\begin{aligned}
R_{l,k} &\equiv \sum_{j \neq p} V_{l,j} V_{k,,j} Y_j \\
&\equiv \sum_{j \neq p} \iint e_l^{ap} \cdot e_j \, ds \; \cdot \iint e_k^{ap} \cdot e_j \, ds \; \cdot Y_j
\end{aligned}
\tag{C.15}
$$

The above matrix equation is thus

$$
\begin{bmatrix}
R_{0,0} & \cdots & R_{0,k} \\
\vdots & \ddots & \\
R_{l,0} & & R_{l,k}
\end{bmatrix}
\cdot
\begin{bmatrix}
1 \\
c_1' \\
\vdots \\
c_k'
\end{bmatrix}
= 0
$$

If we set normalize the $\mathbf{c}$-vector to c_0, then we can make re-arrange things to be a standard non-homogeneous equation (bauwen's idea).

$$
\begin{bmatrix} R_{0,0} \\ \vdots \\ R_{l,0} \end{bmatrix} + \begin{bmatrix} R_{0,1} & \cdots & R_{0,k} \\ \vdots & \ddots & \\ R_{l,1} & & R_{l,k} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ c'_1 \\ \vdots \\ c'_k \end{bmatrix} = 0
$$

$$
\begin{bmatrix} 1 \\ c'_1 \\ \vdots \\ c'_k \end{bmatrix} = \begin{bmatrix} R_{0,1} & \cdots & R_{0,k} \\ \vdots & \ddots & \\ R_{l,1} & & R_{l,k} \end{bmatrix}^{-1} \cdot \begin{bmatrix} R_{0,0} \\ \vdots \\ R_{l,0} \end{bmatrix} \tag{C.16}
$$

Alternatively, we can seperate the dependence on l and k .

$$
\begin{bmatrix} \sum_{j\neq p} V_{0,j} V_{0,j} Y_j & \cdots & \sum_{j\neq p} V_{0,j} V_{k,j} Y_j \\ \vdots & \ddots & \\ \sum_{j\neq p} V_{l,j} V_{0,j} Y_j & & \sum_{j\neq p} V_{l,j} V_{k,j} Y_j \end{bmatrix} = \begin{bmatrix} \sum_{j\neq p} V_{0,j} \sqrt{Y_j} \\ \vdots \\ \sum_{j\neq p} V_{l,j} \sqrt{Y_j} \end{bmatrix} \begin{bmatrix} \sum_{j\neq p} V_{0,j} \sqrt{Y_j} & \cdots & \sum_{j\neq p} V_{k,j} \sqrt{Y_j} \end{bmatrix}
$$

C.4.2. Caution on Re-evaluating Mode Voltages

It is useful to note that integral in C.15 has, in general, three components.

$$
\mathbf{e_j^{ap}} = \hat{\mathbf{x}} \cdot e_j^{ap} + \hat{\mathbf{y}} \cdot e_j^{ap} + \hat{\mathbf{z}} \cdot e_j^{ap}
$$

197

$$\iint_{ap} \mathbf{e_j}^{\mathbf{ap}} \cdot \mathbf{e_j} = \iint_{ap} \hat{\mathbf{x}} \cdot \left(e_j^{ap} \cdot e_j \right) + \iint_{ap} \hat{\mathbf{y}} \cdot \left(e_j^{ap} \cdot e_j \right) + \iint_{ap} \hat{\mathbf{z}} \cdot \left(e_j^{ap} \cdot e_j \right)$$

I point this out because commonly the initial aperture field is simple, and has only one component, $\mathbf{e^{ap}} = \hat{\mathbf{y}} \cdot e^{ap}$. This simplifies the evaluation of the mode-voltages. If one wish's to apply the Raleigh-Ritz to improve upon the trial field, then basis set chosen must be consistent with the condition $\mathbf{e^{ap}} = \hat{\mathbf{y}} \cdot e^{ap}$. Or the mode-voltage integral must be re-evaluated.

C.4.3. General Junction

The

$$Y_{in,p}^{I'} = \frac{\sum_j \left(\iint \mathbf{E^{ap}} \cdot e_j^{II} \, ds \right)^2 Y_{in,j}^{II} + \sum_{i \neq p} \left(\iint \mathbf{E^{ap}} \cdot e_i^{I} \, ds \right)^2 Y_i^{I}}{\left(\iint \mathbf{E^{ap}} \cdot e_p \, ds \right)^2 Y_p}$$

Expressing the aperture field as a sum of weighted basis functions, where the weights are yet to be determined,

$$\mathbf{E^{ap}} = \sum_k c_k e_k^{ap}$$

substituting this into the earlier expression,

$$Y_{in,p}^{I'} = \frac{\sum_j \left(\sum_k \iint c_k e_k^{ap} \cdot e_j^{II} \, ds\right)^2 Y_{in,j}^{II}}{\left(\sum_k \iint c_k e_k^{ap} \cdot e_p \, ds\right)^2 Y_p} + \frac{\sum_{i \neq p} \left(\sum_k \iint c_k e_k^{ap} \cdot e_i^{I} \, ds\right)^2 Y_i^{I}}{\left(\sum_k \iint c_k e_k^{ap} \cdot e_p \, ds\right)^2 Y_p}$$

D. Verification of Variational Method

D.1. Introduction

The following is an application of the variation expression for junction admittance for the purpose of calculating the dominant mode scattering parameters off a translation offset between two rectangular waveguides. The geometry of the cross section of junction is shown below. The model used to calculate the admittance is based on the variational expression originally formulated by Harrington, Collins, and Marcuvitz, which is re-derived in the previous chapter. This expression relies on an assumed field and thus, is only an approximation. The model also has the benefit that it provides good intuition of the physics involved.
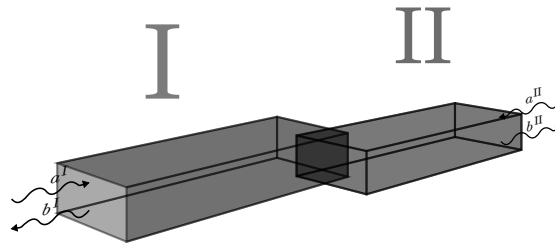
Figure D.1.: Geometry of Cross section
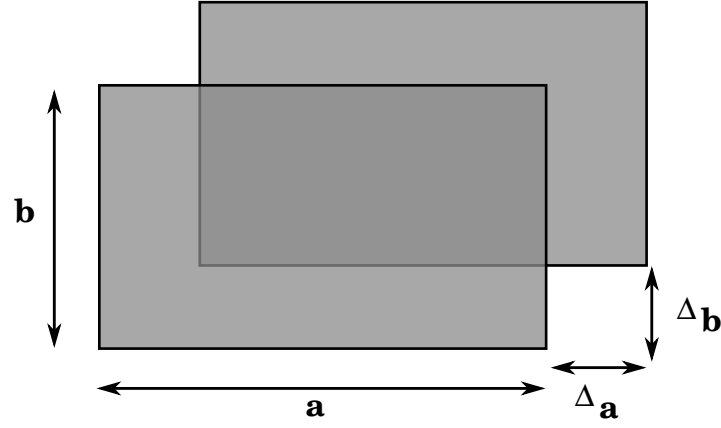
Figure D.2.: Geometry of Cross section

D.2. Outline

As described in the previous chapter, the junction admittance can be analyzed either by assuming knowledge of the tangential field, or the tangential currents. This section is based on assumed knowledge of the electric field, so we want to evaluate this expression,

$$
Y_{in,p}^{'\,I} = \frac{\sum_j V_j^2 Y_j Y_{in,j}^{'\,II} - \sum_{i\neq p} V_i^2 Y_i Y_{in,i}^{'\,I}}{V_p^2 Y_p^{I}}
$$

The outline for the analysis is as follows:

1. Make educated guess at the functional form of aperture field; E^{ap}

2. Given E^{ap} from step 1, determine mode-voltages (aka coupling coefficients) for each waveguide; V^I, V^{II}.

3. Account for terminations, by calculating the waveguide input admittances, as seen from junction.

4. Evaluate variational expression, to get normalized junction admittance, Y'_{in} , for dominant mode.

5. Calculate reflection coefficient for dominant mode, $S_{11} = \frac{1-Y'_{in}}{1+Y'_{in}}$

The only part of the analysis which changes with different junction geometries is the evaluation of mode voltages. The first junction will be worked out in detail, while the following calculations will just be evaluation of the mode voltages.

D.2.1. Capacitive Junction

First we will examine the special case, of $\Delta_a = 0$, so there is only a step in the $'b'$ dimension. Due to the modes this kind of offset excites, it produces an admittance which is capacitive. The cross section of the junction as seen by an input waveguide is shown below.
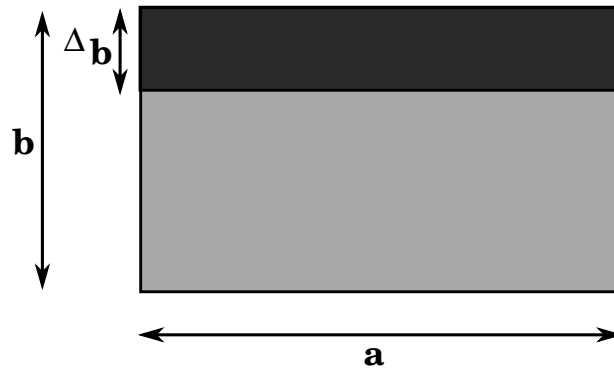
Figure D.3.: Cross section as seen from input waveguide

D.2.1.1. Assumed Field

The assumed aperture field is the dominant mode of the common cross-section between the two guides.

$$\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{a}x\right) & 0 < y < b - \Delta_b \\ \\ 0 & else \end{cases}$$

D.2.1.2. Mode Voltages

Due to the symmetry of this problem, both the eigen-functions and mode-voltages-squared for each guide are the same.

$$\left(V^I\right)^2 = \left(V^{II}\right)^2$$

Because we are assuming only a $\hat{y}$-component for the aperture field, it will only couple to the $\hat{y}$-component of the normalized eigen-functions in either guide. There freedom in which set of eigenfunctions we choose. T-to-x ,T-to-y, and T-to-z will all produce the same end results. The advantage of one set over the other is some cases the integrals may be more tractable. I will choose T-to-z.

The normalized TE-z and TM-z eigenfunctions for rectangular waveguide are given in the appendix. Their y-components are,

$$\mathbf{e_{(m,n,l)}} \cdot \hat{y} = \begin{cases} \dfrac{-\sqrt{\epsilon_m \epsilon_n}}{a} \cdot \dfrac{m}{\sqrt{m^2 \frac{b}{a} + n^2 \frac{a}{b}}} \cdot sin\left(\frac{m\pi}{a}x\right) cos\left(\frac{n\pi}{b}y\right) & l =' te'; \quad (m,n) \neq (0,0) \\[4mm] \dfrac{-2}{b} \cdot \dfrac{n}{\sqrt{m^2 \frac{b}{a} + n^2 \frac{a}{b}}} \cdot sin\left(\frac{m\pi}{a}x\right) cos\left(\frac{n\pi}{b}y\right) & l =' tm'; \quad m \neq 0, n \neq 0 \end{cases}$$

Where ϵ_i is called *neuman's number*, and is given by; $\epsilon_i \equiv 2 - \delta\left(i\right)$.

It is interesting to note that the functional forms for each mode are identical, and that they only differ by a scaling factor in front. This scaling factor is not a function of the independent variables of the cross-section (x and y), so it wont effect the integrals. Furthermore, its the same expression for all rectangular waveguides, so to make the expressions more readable I will represent this scaling factor with $f\left(..\right)$.

$$\mathbf{e_{(m,n,l)}} = \underbrace{f\left(a,b,m,n,l\right)}_{\text{independent on x,y}} \cdot \underbrace{sin\left(\frac{m\pi}{a}x\right) cos\left(\frac{n\pi}{b}y\right)}_{\text{dependent on x,y}}$$

The evaluation of the mode voltages, which represent the coupling of the assumed field to a given waveguide mode. Introducing the variable, $d = b - \Delta_b$, for concision, the mode voltage is found by,

$$
\begin{aligned}
\mathbf{V}_{(\mathbf{m},\mathbf{n},\mathbf{l})} \;&=\; \int\limits_{0}^{d} dy \int\limits_{0}^{a} dx, \; \mathbf{E^{ap}} \cdot \mathbf{e_{m,n,l}} \\[2mm]
&=\; f_{(..)} \int\limits_{0}^{d} dy \int\limits_{0}^{a} dx, \; sin\left(\frac{\pi}{a}x\right) \cdot sin\left(\frac{m\pi}{a}x\right) cos\left(\frac{n\pi}{b}y\right) \\[2mm]
&=\; f_{(..)} \int\limits_{0}^{d} dy \int\limits_{0}^{a} dx, \; \delta\left(m-1\right) sin^{2}\left(\frac{\pi}{a}x\right) cos\left(\frac{n\pi}{b}y\right) \\[2mm]
&=\; f_{(..)}\frac{a}{2}\delta\left(m-1\right) \cdot \int\limits_{0}^{d} dy, \, cos\left(\frac{n\pi}{b}y\right) \\[2mm]
&=\; f_{(..)}\frac{a}{2}\delta\left(m-1\right) \cdot \frac{b}{n\pi} sin\left(\frac{n\pi}{b}d\right)
\end{aligned}
$$

This is correct, but notice that the evaluation at $n = 0$ requires the use of $l'\,hopital's$ rule. Also, if you wanted to program this function, it would require an extra if/else block. This problem can be avoided by recognizing we re-write the expression in terms of the sinc function[1].

$$
sinc\left(x\right) \equiv \frac{sin\left(\pi x\right)}{\pi x}
$$

Re-writing the final expression for the coupling.

---

[1]note: This is the *normalized* sinc function. Make sure you know if you are using a *normalized* or *un-normalized* sinc function.

$$
\begin{aligned}
\mathbf{V_{(m,n,l)}} &= f_{(..)}\frac{a}{2}\delta\left(m-1\right)\cdot\frac{b}{n\pi}sin\left(\frac{n\pi}{b}d\right) \\
&= f_{(..)}\frac{a}{2}\delta\left(m-1\right)\cdot d\frac{sin\left(\frac{n\pi}{b}d\right)}{\frac{n\pi}{b}d} \\
&= f_{(..)}\frac{a}{2}\delta\left(m-1\right)\cdot d\ sinc\left(\frac{d}{b}n\right)
\end{aligned}
$$

The subtle removal of the π was because I have defined sinc as the *normalized* sinc function. Re-inserting the scaling factor, yields the complete expression for the mode-voltage

$$
\mathbf{V_{(m,n,l)}} = \begin{cases}
f_{(..)}\cdot\delta\left(m-1\right)\cdot d\ sinc\left(\frac{d}{b}n\right) & l =' te';\quad (m,n)\neq(0,0) \\[2ex]
f_{(..)}\cdot\delta\left(m-1\right)\cdot d\ sinc\left(\frac{d}{b}n\right) & l =' tm';\quad m\neq 0, n\neq 0
\end{cases}
$$

A plot of the mode-voltages squared, for $m = 1$, is shown below. This gives an idea of how many modes need to be included in summation before convergence is reached. I have plotted the function vs a continuous variable 'n' solely to illustrate its functional form. In reality this function is only valid at integer multiples of n-indices's. This is the cause of the apparent discontinuity of the $'te'$ mode at $n = 0$.
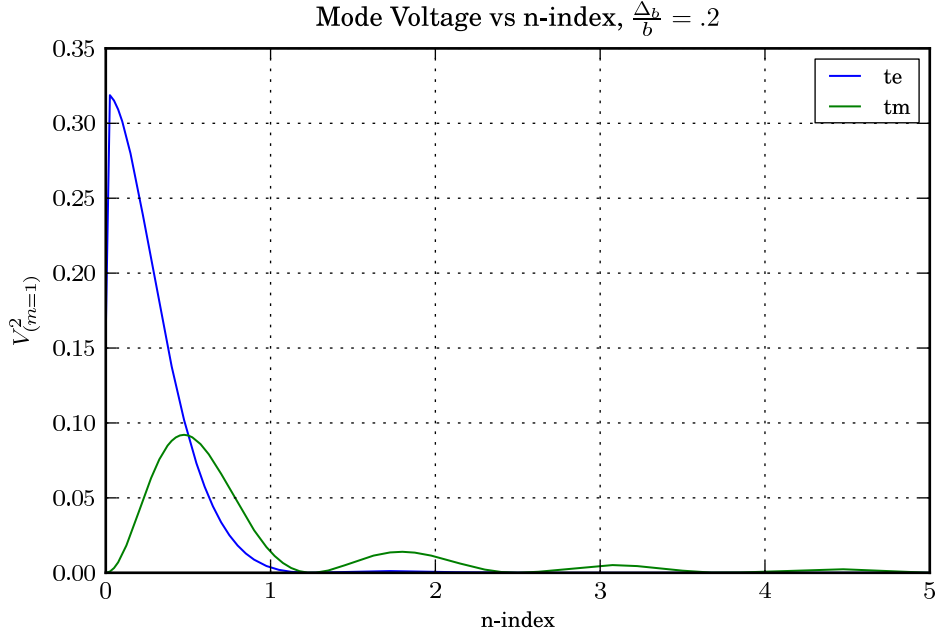
Figure D.4.: Capacitive Junction Mode Voltages Vs n-index ,evaluated at m=0

D.2.1.3. Waveguide Admittances

The characteristic admittances in rectangular waveguide are given by

$$
Y_{0,(m,n,l)} = \begin{cases} \frac{-\omega\epsilon}{k_z} & l =' tm' \\[2ex] \frac{k_z}{\omega\mu} & l =' te' \end{cases}
$$

The waveguide input admittances depend on the terminations. Any termination which can be faithfully represented by an input admittance can be modeled by this technique. For simplicity we will assume a matched termination at both waveguides. Other For matched conditions in region I,

$$
Y'^{I}_{in,i\neq p} = \left.\frac{(a_i - b_i)^{I}}{(a_i + b_i)}\right|_{i\neq p} = -1
$$

For matched conditions in region II,

$$Y_{in,j}^{'II} = \frac{(b_j - a_j)^{II}}{(b_j + a_j)} = 1$$

See my other paper for a derivation of this.

D.2.1.4. Junction Admittance

At each frequency of interest, the junction admittance is calculated by

$$\mathbf{Y_{in,p}^{I'}} = \frac{\sum_j \left(\mathbf{V_j^{II}}\right)^2 Y_j^{II} + \sum_{i \neq p} \left(\mathbf{V_i^{I}}\right)^2 Y_i^{I}}{\left(\mathbf{V_p^{I}}\right)^2 Y_p}$$

Where the indices's i and j are interpreted as all possible modes, but will in practice be truncated. Acceptable truncation values can be systematically determined by defining some kind of convergence metric, such as maximum complex difference of successive admittance values for increasing number of modes. This is discused in detail in the implementation chapter.

D.2.1.5. Results

The reflection coefficient of a given junction admittance is calculated by

$$S_{11,p} = \frac{1 - Y_{in,p}^{'}}{1 + Y_{in,p}^{'}}$$

Below are the results of the above analysis evaluated for WR1.5 waveguide junction with a offset of $\Delta_b = \frac{a}{10}$, using 11 modes for each dimension ($M, N = 11, 11$). Note the phase response illustrates that this junction looks capacitive.

Figure D.5.:

Offset Waveguide Return Loss Magnitude: $\frac{\Delta_a}{a} = 0.0, \frac{\Delta_b}{a} = 0.1$

Figure D.6.:

Offset Waveguide Return Loss Phase: $\frac{\Delta_a}{a} = 0.0, \frac{\Delta_b}{a} = 0.1$
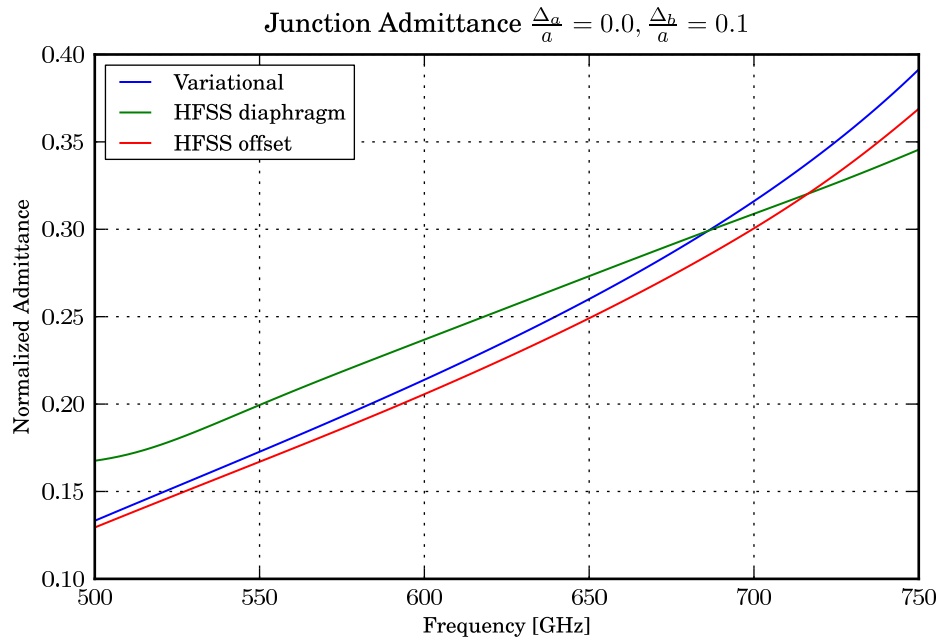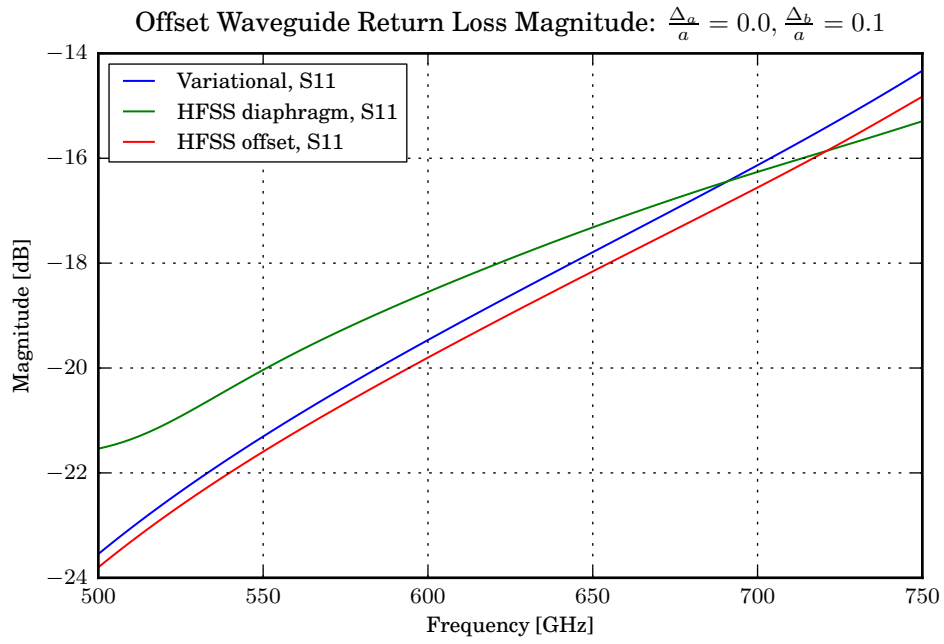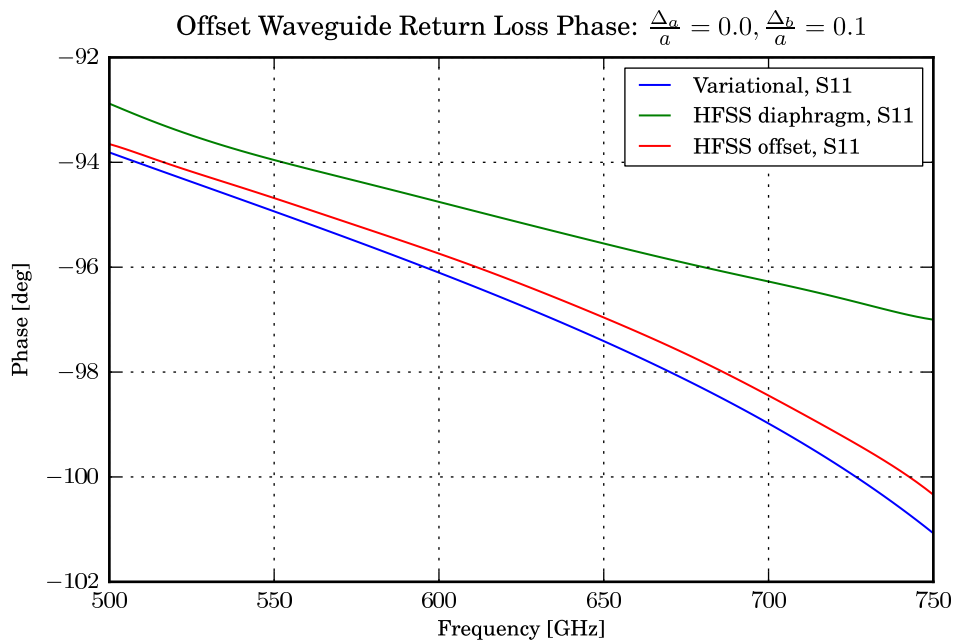
Figure D.7.:

D.2.2. Inductive Offset

Next, we will examine the special case, of $\Delta_b = 0$, so there is only a step in the $'a'$ dimension. Due to the modes this kind of offset excites, it produces an admittance which is inductive. An illustration of the cross section of the junction as seen by an input waveguide is shown below.
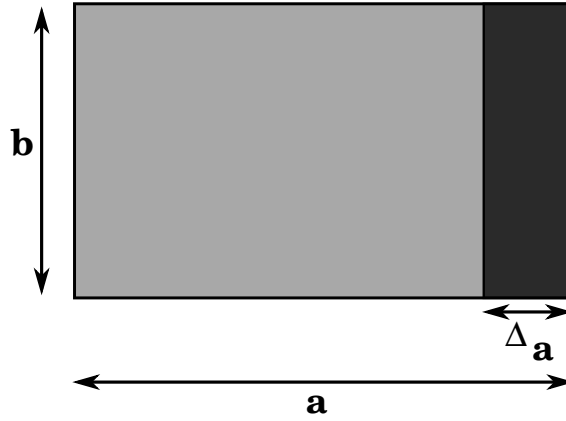
Figure D.8.: Cross section as seen from input waveguide

D.2.2.1. Assumed Field

Assuming the aperture field is the dominant mode in the common cross-section,

$$
\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{(a-\Delta_a)}x\right) & 0 < x < a - \Delta_a, \, 0 < y < b \\ 0 & else \end{cases}
$$

Note this is only the dominant mode as long as $a > b$. The analysis of the inductive offset is carried out in similar manner to the capacitive offset, with the difference

that the aperture field, and cross-section have changed.

D.2.2.2. Mode Voltages

The mode voltages are found in an analogous way to the capacitive junction, with a change in the integral bounds. Introducing variable, $c = a - \Delta_a$

$$
\begin{aligned}
\mathbf{V_{(m,n,l)}} &= \int\limits_0^c dx \int\limits_0^b dy, \ \mathbf{E^{ap}} \cdot \mathbf{e_{m,n,l}} \\
&= f_{(..)} \int\limits_0^c dx \int\limits_0^b dy, \ cos\left(\frac{n\pi}{b}y\right) sin\left(\frac{\pi}{c}x\right) sin\left(\frac{m\pi}{a}x\right) \\
&= f_{(..)} b\, \delta(n) \int\limits_0^c dx, \ sin\left(\frac{\pi}{c}x\right) sin\left(\frac{m\pi}{a}x\right)
\end{aligned}
$$

This last integral is a little tricky[2], but making use of the mixing trig identity[3],

$$
sin\left(a\right) sin\left(b\right) = \frac{1}{2}\left[cos\left(a - b\right) - cos\left(a + b\right)\right]
$$

---

[2]If you are lazy, or want to check your work, you can put this directly into wolfram alpha; http://www.wolframalpha.com. Just enter: Integral(sin(m*pi*x/a)*sin(pi*x/c),(x,0,c))

[3]http://en.wikipedia.org/wiki/List_of_trigonometric_identities

$$= \int_0^c dx, \, sin\left(\frac{\pi}{c}x\right) sin\left(\frac{m\pi}{a}x\right)$$

$$= \int_0^c dx, \, \frac{1}{2}\left(cos\left[\left(\frac{\pi}{c} - \frac{m\pi}{a}\right)x\right] - cos\left[\left(\frac{\pi}{c} + \frac{m\pi}{a}\right)x\right]\right)$$

$$= \left(\frac{sin\left[\left(\frac{\pi}{c} - \frac{m\pi}{a}\right)x\right]}{2\left(\frac{\pi}{c} - \frac{m\pi}{a}\right)}\right)\Bigg|_0^c - \left(\frac{sin\left[\left(\frac{\pi}{c} + \frac{m\pi}{a}\right)x\right]}{2\left(\frac{\pi}{c} + \frac{m\pi}{a}\right)}\right)\Bigg|_0^c$$

$$= \frac{sin\left[\left(\frac{\pi}{c} - \frac{m\pi}{a}\right)c\right]}{2\left(\frac{\pi}{c} - \frac{m\pi}{a}\right)} - \frac{sin\left[\left(\frac{\pi}{c} + \frac{m\pi}{a}\right)c\right]}{2\left(\frac{\pi}{c} + \frac{m\pi}{a}\right)}$$

$$= \frac{c}{2} \cdot \left[sinc\left(1 - \frac{c}{a}m\right) - sinc\left(1 + \frac{c}{a}m\right)\right]$$

Including the earlier found $'n'$-dependence, we have a final expression for the mode-voltages.

$$\mathbf{V_{(m,n,l)}} = \begin{cases} f_{(..)} \cdot \delta(n)\frac{bc}{2} \cdot \left[sinc\left(1 - \frac{c}{a}m\right) - sinc\left(1 + \frac{c}{a}m\right)\right] & l =' te'; \quad (m,n) \neq (0,0) \\ \\ 0 & l =' tm'; \end{cases}$$

Only the $n = 0$ modes are excited, which means the TM mode voltages will all be zero, because TM modes require $m \neq 0$ and $n \neq 0$. Shown below the TE mode voltage vs m-index, for an offset of $\frac{\Delta_a}{a} = \frac{1}{10}$. The dominant TE10 mode is most strongly coupled to, which makes sense.
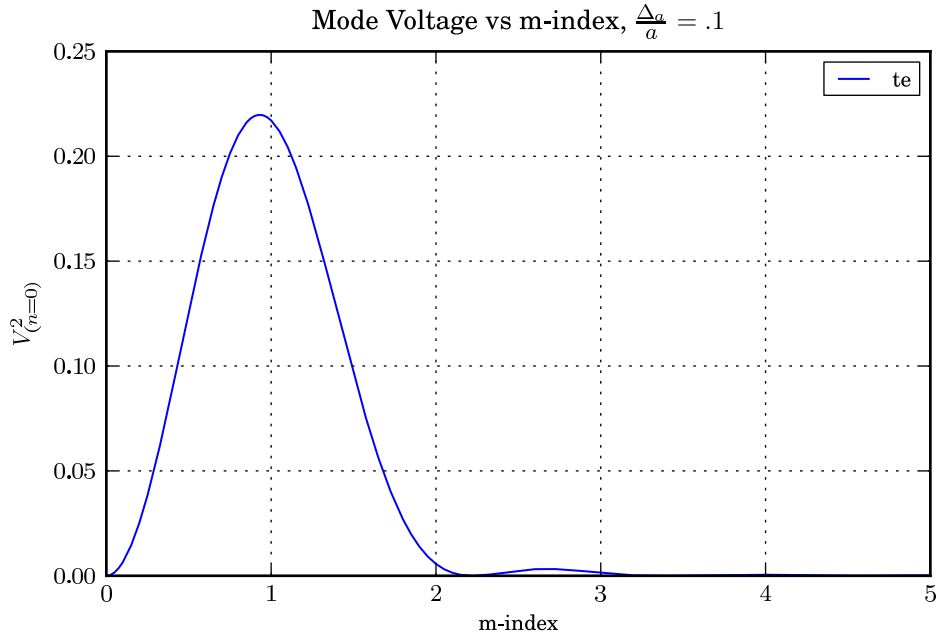
Figure D.9.:

D.2.2.3. Results

Below are the results of the above analysis evaluated for WR1.5 waveguide junction with a offset of $\Delta_a = \frac{a}{10}$, using 11 modes for each dimension ($M, N = 11, 11$). For validation, the results of a HFSS simulation of the same geometry is shown. Note the phase response illustrates that this junction looks inductive, as expected.
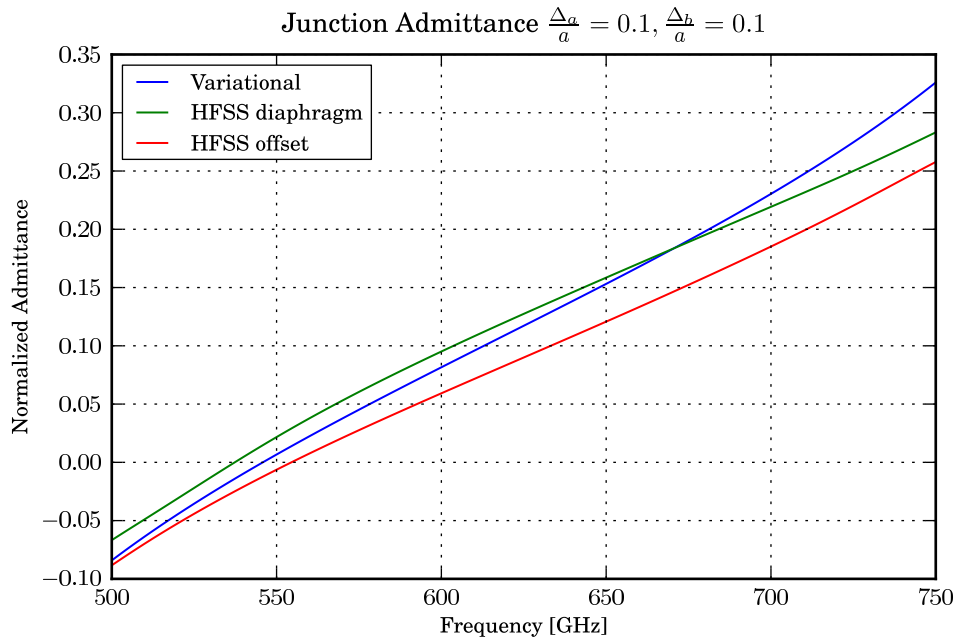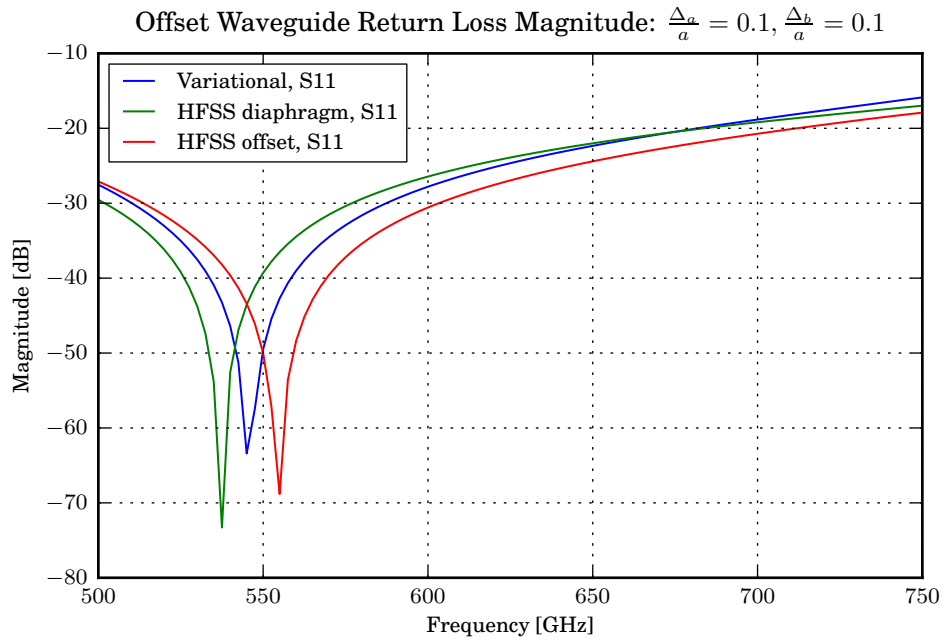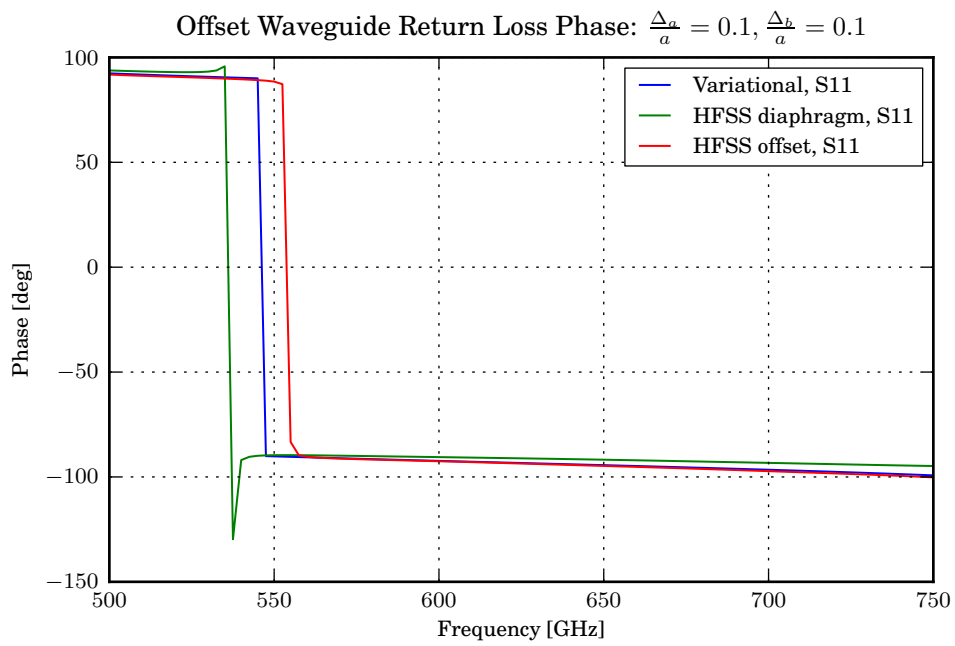
Figure D.10.:

Figure D.11.:

Offset Waveguide Return Loss Phase: $\frac{\Delta_a}{a} = 0.1, \frac{\Delta_b}{a} = 0.0$
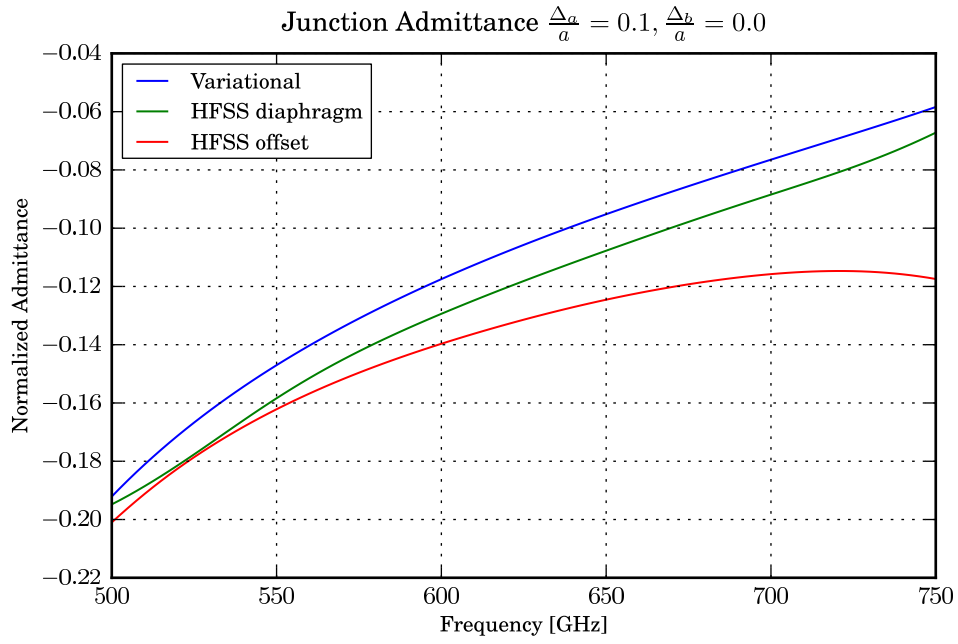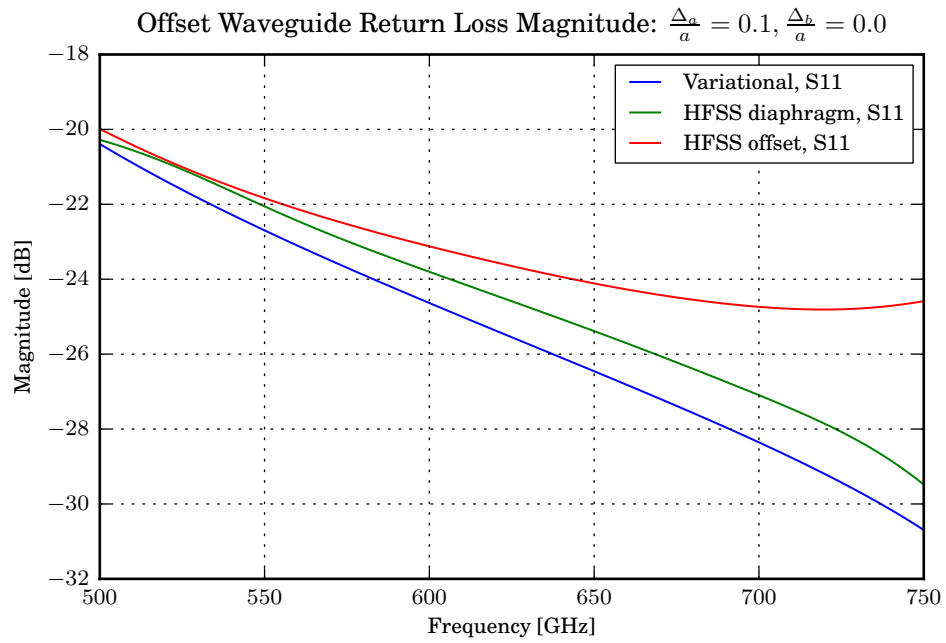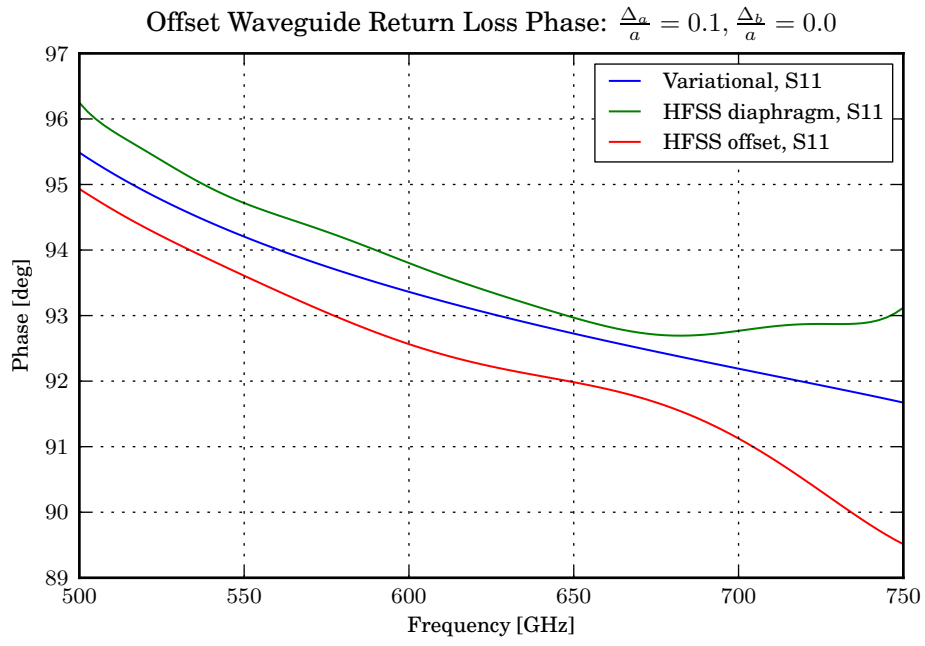
Figure D.12.:

D.2.3. Resonant Offset

Finally, we will examine the general case, of $\Delta_b \neq 0, \Delta_a \neq 0$. This junction excites both inductive and capacitive modes, and thus acts like a resonant LC circuit. An illustration of the cross section of the junction as seen by an input waveguide is shown below.
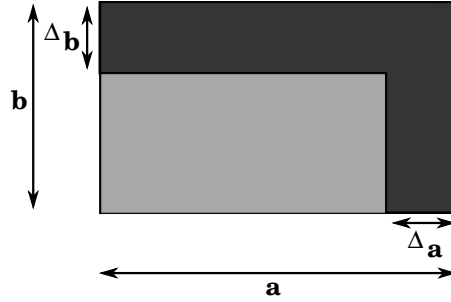
Figure D.13.: Cross section as seen from input waveguide

D.2.3.1. Assumed Field

Assuming the aperture field is the dominant mode in the common cross-section,

$$
\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{(a-\Delta_a)}x\right) & 0 < x < a - \Delta_a, \quad 0 < y < b - \Delta_b \\ 0 & else \end{cases}
$$

D.2.3.2. Mode Voltages

Making use of the solutions to the integrals we found in the evaluation of the capacitive and inductive offsets, we can immediately write down the answer,

$$
\begin{aligned}
\mathbf{V_{(m,n,l)}} &= \int_0^{a-\Delta_a} dx \int_0^{b-\Delta_b} dy, \ \mathbf{E^{ap}} \cdot \mathbf{e_{m,n,l}} \\
&= f_{(..)} \int_0^{a-\Delta_a} dx \int_0^{b-\Delta_b} dy, \ cos\left(\frac{n\pi}{b}y\right) sin\left(\frac{\pi}{a}x\right) sin\left(\frac{m\pi}{a}x\right) \\
&= = f_{(..)} \frac{cd}{2}\left[sinc\left(\frac{d}{b}n\right)\right] \cdot \left[sinc\left(1 - \frac{c}{a}m\right) - sinc\left(1 + \frac{c}{a}m\right)\right]
\end{aligned}
$$

Explicitly writing the eigenfunction scaling factors,

$$
\mathbf{V_{(m,n,l)}} = \begin{cases} f_{(..)} \cdot \frac{cd}{2}\left[sinc\left(\frac{d}{b}n\right)\right] \cdot \left[sinc\left(1-\frac{c}{a}m\right) - sinc\left(1+\frac{c}{a}m\right)\right] & l =' te'; \quad (m,n) \neq (0,0) \\[2ex] f_{(..)} \cdot \frac{cd}{2}\left[sinc\left(\frac{d}{b}n\right)\right] \cdot \left[sinc\left(1-\frac{c}{a}m\right) - sinc\left(1+\frac{c}{a}m\right)\right] & l =' tm'; \quad m \neq 0, n \neq 0 \end{cases}
$$

Because this junction couples many modes in m and n , in general, illustrating the nature of the coupling requires a 2D surface. Below is colored image of the coupling matrix to each mode-type sampled at integer values of m and n.
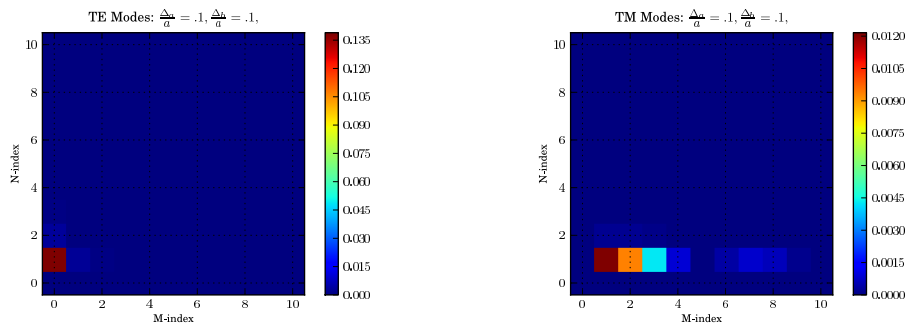
Figure D.14.: Coupling Matrices for $\frac{\Delta_a}{a} = \frac{1}{10}, \frac{\Delta_b}{a} = \frac{1}{10}$

D.2.3.3. Results

Below are the results of the above analysis evaluated for WR1.5 waveguide junction with a offset of $\Delta_a = \frac{a}{10}, \Delta_b = \frac{a}{10}$, using 11 modes for each dimension ($M, N = 11, 11$). For validation, the results of a HFSS simulation of the same geometry is shown. Note the phase response illustrates that this junction looks inductive, as expected.

Figure D.15.:

Offset Waveguide Return Loss Magnitude: $\frac{\Delta_a}{a} = 0.1$, $\frac{\Delta_b}{a} = 0.1$

Figure D.16.:

Offset Waveguide Return Loss Phase: $\frac{\Delta_a}{a} = 0.1$, $\frac{\Delta_b}{a} = 0.1$

Figure D.17.:

D.2.4. General Case: Dimensional change with offset

From the first few examples it should start to be clear that each side of the junction can be dealt with independently from the other. The only requirement is that assumed field must be consistent between the two regions. This means one can solve a more general aperture geometry and then tailor it to suite a variety of discontinuities. A cross-sectional geometry which can provide solutions for all of the previous problems is illustrated in the figure below.

The larger guide is of dimensions AxB, the aperture is axb. The lower left corner of the aperture is offset from the larger's, by an aribitrary amount (x_0, y_0).
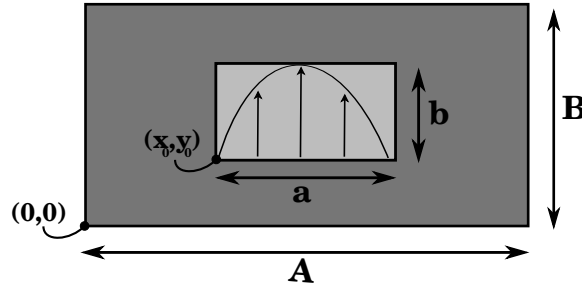
Figure D.18.: Step Down

One may imagine this aperture field is due to a smaller guide radiating into a larger guide. The assumed aperture field will be the dominant mode of the smaller guide.

$$
\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) & x_0 < x < x_0 + a \quad y_0 < y < y_0 + b \\ 0 & else \end{cases}
$$

The mode volatge will be

$$\mathbf{V^I}_{(m,n,l)} = \int\limits_{x_0}^{x_0+a} dx \int\limits_{y_0}^{y_0+b} dy, \ \mathbf{E^{ap}} \cdot \mathbf{e_{m,n,l}}$$

$$= f(..) \cdot \int\limits_{x_0}^{x_0+a} dx \int\limits_{y_0}^{y_0+b} dy, \ sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) \cdot sin\left(\frac{m\pi}{A}x\right) cos\left(\frac{n\pi}{B}y\right)$$

This integral is serperable into the product of two integrals. Each integral also represents a single dimension, meaning a step down in width is handled by the x-dependent part, while a step in height is handeled by the y-dependent part.

$$\mathbf{V^I}_{(m,n,l)} = f(..) \cdot \left(\underbrace{\int\limits_{x_0}^{x_0+a} dx, \ sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) \cdot sin\left(\frac{m\pi}{A}x\right)}_{\text{width step}} \cdot \underbrace{\int\limits_{y_0}^{y_0+b} dy, \ cos\left(\frac{n\pi}{B}y\right)}_{\text{height step}} \right)$$

First the height step integral,

$$= \int\limits_{y_0}^{y_0+b} dy, \ cos\left(\frac{n\pi}{B}y\right)$$

$$= \frac{B}{n\pi}sin\left(\frac{n\pi}{B}y\right)\Bigg|_{y_0}^{|y_0+b}$$

$$= \frac{B}{n\pi}sin\left(\frac{n\pi}{B}(y_0+b)\right) - \frac{B}{n\pi}sin\left(\frac{n\pi}{B}y_0\right)$$

$$= (y_0+b) \cdot sinc\left(\frac{n}{B}[y_0+b]\right) - y_0 \cdot sinc\left(\frac{n}{B}y_0\right)$$

Second the width step part,

$$
= \int_{x_0}^{x_0+a} dx, \, sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) \cdot sin\left(\frac{m\pi}{A}x\right)
$$

$$
= \int_{x_0}^{x_0+a} dx, \, cos\left(\left[\frac{\pi}{a} - \frac{m\pi}{A}\right]x + \frac{\pi}{a}x_0\right) + \int_{x_0}^{x_0+a} dx, \, cos\left(\left[\frac{\pi}{a} + \frac{m\pi}{A}\right]x - \frac{\pi}{a}x_0\right)
$$

$$
= \frac{1}{2}\left(\frac{1}{\left(\frac{m\pi}{A} - \frac{\pi}{a}\right)}\left[sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0 + a] + \frac{\pi}{a}x_0\right) - sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0] + \frac{\pi}{a}x_0\right) \right] - \right.
$$

$$
\left. \frac{1}{\left(\frac{m\pi}{A} + \frac{\pi}{a}\right)}\left[sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0 + a] - \frac{\pi}{a}x_0\right) - sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0] - \frac{\pi}{a}x_0\right) \right] \right)
$$

This form has a singularity at $\frac{A}{a} = \text{int}$, which can be avoided either by perturbing one of the widths so that the ration is never an integer, or by using l'hopital. The perturbation is easier and more efficient to implement computationallly.

In summary, the total coupling of an offset rectangular aperture in a rectangular waveguide is,

$$
\mathbf{V^I}_{(m,n,l)} = f_{(..)} \cdot \left[(y_0 + b) \cdot sinc\left(\frac{n}{B}[y_0 + b]\right) - y_0 \cdot sinc\left(\frac{n}{B}y_0\right) \cdot \right.
$$

$$
\frac{1}{2}\left(\frac{1}{\left(\frac{m\pi}{A} - \frac{\pi}{a}\right)}\left[sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0 + a] + \frac{\pi}{a}x_0\right) - sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0] + \frac{\pi}{a}x_0\right) \right] - \right.
$$

$$
\left. \left. \frac{1}{\left(\frac{m\pi}{A} + \frac{\pi}{a}\right)}\left[sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0 + a] - \frac{\pi}{a}x_0\right) - sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0] - \frac{\pi}{a}x_0\right) \right] \right] \right.
$$

Next we will use this result to solve a roated waveguide discontinuity.

D.2.5. Rotated Waveguide

The rotated waveguide, although initially measured on accident, can be used as a standard as well. Due to the orthogonality of dominant modes in each region there is no propagating mode in the ouput region. This means the response is solely determined by the evanscent fields in the second region. If the second region is shorted close to the junction, the evanscent fields will be reflected with a different phase than if the region was matched. One advantage of using this as a standard is that i think it will be less sensitive to translation misalignment. This has yet to be rigorously proved.
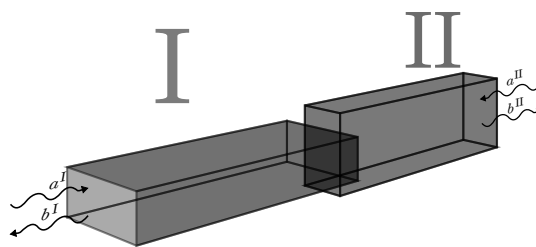
Figure D.19.: Rotated waveguide Junction

We will first solve this for a perfect alignment, meaning both guides are share common center, and then explore how sensitive it is to misalignment, to see if we need to formulate the general case.
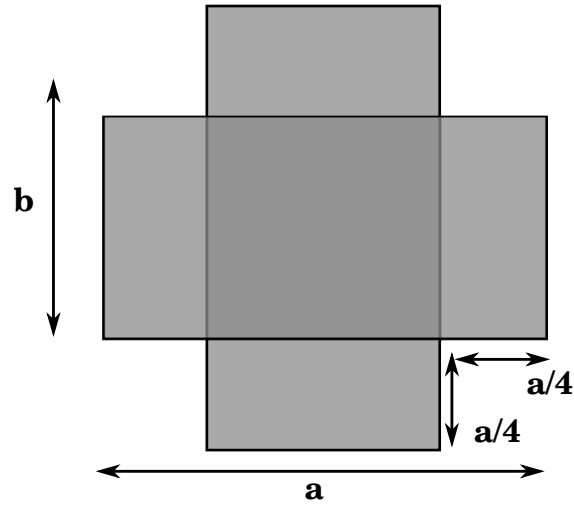
Figure D.20.: Rotated waveguide cross-section

D.2.5.1. Assumed Field

Assuming the aperture field is the dominant mode in the common cross-section,

$$
\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) & \frac{a}{4} < x < \frac{3a}{4} \quad 0 < y < b \\ 0 & else \end{cases}
$$

Due to the asymetry, between the aperture field and the two regions, the mode-voltages will be different for each region.

D.2.5.2. Mode Voltages: Region I (input)
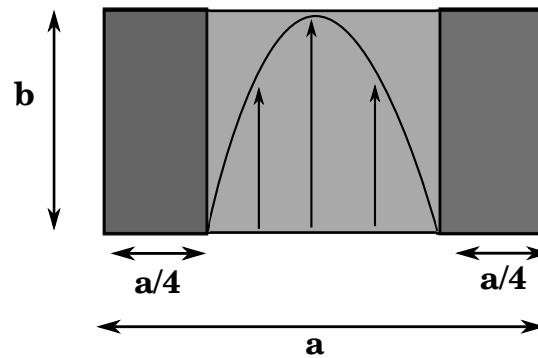
Figure D.21.: Cross section as seen from input waveguide

Instead of re-doing the integrals, we can just tailor this problem to fit into the general problem we just solved of 'change is guide size with arbitrary offset'.

$$\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) & x_0 < x < x_0 + a \quad y_0 < y < y_0 + b \\ 0 & else \end{cases}$$

Fitting into this notation, the guide is of size axb and the aperture is of size $\frac{a}{2}$xb with an offset of $\left(\frac{a}{4}, 0\right)$. To be clear, the relation between the previous more general problem and this problem are shown in the table below.

General	Currently
A	a
B	b
a	$\frac{a}{2}$
b	b
x_0, y_0	$\frac{a}{4}, 0$

Skipping straight to the result,

$$
\mathbf{V^I}_{(m,n,l)} = f_{(..)} \cdot \left[\left[(y_0 + b) \cdot sinc\left(\frac{n}{B} [y_0 + b] \right) - y_0 \cdot sinc\left(\frac{n}{B} y_0 \right) \right] \cdot \right.
$$

$$
\frac{1}{2} \left(\frac{1}{\left(\frac{m\pi}{A} - \frac{\pi}{a} \right)} \left[sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a} \right] [x_0 + a] + \frac{\pi}{a} x_0 \right) - sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a} \right] [x_0] + \frac{\pi}{a} x_0 \right) \right] -
$$

$$
\left. \frac{1}{\left(\frac{m\pi}{A} + \frac{\pi}{a} \right)} \left[sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a} \right] [x_0 + a] - \frac{\pi}{a} x_0 \right) - sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a} \right] [x_0] - \frac{\pi}{a} x_0 \right) \right] \right]
$$

Substituting in our variables and making a lot of tedius simplificiations,

$$
\mathbf{V^I}_{(m,n,l)} = f_{(..)} \cdot \left[b \cdot sinc\left(n \right) \cdot \frac{-a}{\pi} \frac{m}{m^2 + 4} \left(sin\left[\frac{3\pi}{4} m \right] + sin\left[\frac{\pi}{4} m \right] \right) \right]
$$

D.2.5.3. Mode Voltages: Region II (output)
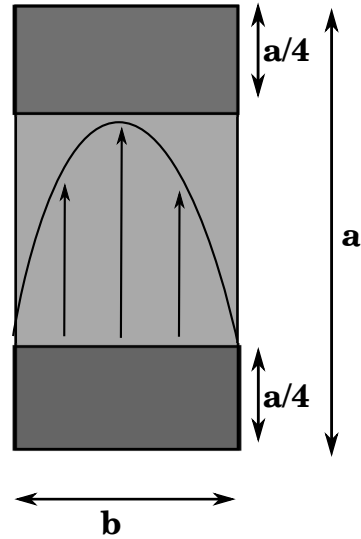
Figure D.22.: Cross section as seen from output waveguide

Region II (the output waveguide) can be analyzed as is, or in rotated coordinates. The advantage to rotating cordinates is that it more closely resembles what we have already analyzed. Tailoring the general result to fit this discontinuity, we have the assumed field

$$
\mathbf{E^{ap}} = \begin{cases} \hat{y} \cdot sin\left(\frac{\pi}{a}x - \frac{\pi}{a}x_0\right) & x_0 < x < x_0 + a \quad y_0 < y < y_0 + b \\ \\ 0 & else \end{cases}
$$

The guide is of size $\frac{a}{2}$x$2b$ and the aperture is of size $\frac{a}{2}$xb with an offset of $\left(0, \frac{b}{2}\right)$. To be clear, the relation between the previous more general problem and this problem are shown in the table below.

General	Currently
A	$\frac{a}{2}$
B	$2b$
a	$\frac{a}{2}$
b	b
x_0, y_0	$0, \frac{b}{2}$

Skipping straight to the result,

$$
\begin{aligned}
\mathbf{V^I}_{(m,n,l)} = {}& f_{(..)} \cdot \left[(y_0 + b) \cdot sinc\left(\frac{n}{B}[y_0 + b]\right) - y_0 \cdot sinc\left(\frac{n}{B}y_0\right) \cdot \right. \\
& \frac{1}{2}\left(\frac{1}{\left(\frac{m\pi}{A} - \frac{\pi}{a}\right)} \left[sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0 + a] + \frac{\pi}{a}x_0\right) - sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0] + \frac{\pi}{a}x_0\right) \right] - \right. \\
& \left.\left. \frac{1}{\left(\frac{m\pi}{A} + \frac{\pi}{a}\right)} \left[sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0 + a] - \frac{\pi}{a}x_0\right) - sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0] - \frac{\pi}{a}x_0\right) \right] \right] \right.
\end{aligned}
$$

Substituting in our variables and making a lot of tedius simplificiations,

$$
\begin{aligned}
\mathbf{V^I_{(m,n,l)}} \;=\; & f_{(..)} \cdot \left[(y_0 + b) \cdot sinc\left(\frac{n}{B}[y_0 + b] \right) - y_0 \cdot sinc\left(\frac{n}{B}y_0 \right) \cdot \right. \\
& \frac{1}{2}\left(\frac{1}{\left(\frac{m\pi}{A} - \frac{\pi}{a}\right)} \left[sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0 + a] + \frac{\pi}{a}x_0 \right) - sin\left(\left[\frac{m\pi}{A} - \frac{\pi}{a}\right][x_0] + \frac{\pi}{a}x_0 \right) \right] - \right. \\
& \left. \left. \frac{1}{\left(\frac{m\pi}{A} + \frac{\pi}{a}\right)} \left[sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0 + a] - \frac{\pi}{a}x_0 \right) - sin\left(\left[\frac{m\pi}{A} + \frac{\pi}{a}\right][x_0] - \frac{\pi}{a}x_0 \right) \right] \right] \right.
\end{aligned}
$$

D.2.5.4. Mode Voltages: Region II Alternative Integrals, (output)

The aperture assumed for Region I is in Region II,

$$
\mathbf{E^{ap}} =
\begin{cases}
\hat{y} \cdot sin\left(\frac{2\pi}{a}x \right) & 0 < x < b \quad \frac{a}{4} < y < \frac{3a}{4} \\
\\
0 & else
\end{cases}
$$

Rotating the coordinates this becomes,
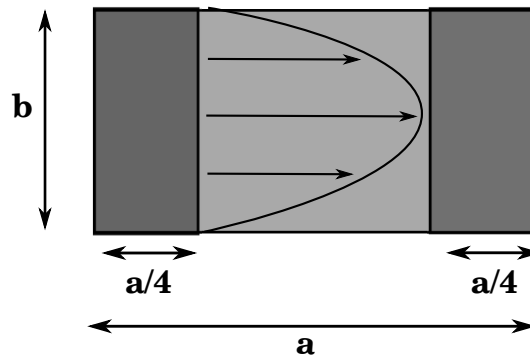
Figure D.23.: Rotated cross section as seen from output waveguide

$$
\mathbf{E^{ap}} = \begin{cases} \hat{x} \cdot sin\left(\frac{2\pi}{a}y\right) & \frac{a}{4} < x < \frac{3a}{4} \quad 0 < y < b \\ \\ 0 & else \end{cases}
$$

Evaluating the mode voltages

$$
\begin{aligned}
\mathbf{V^{II}}_{\mathbf{(m,n,l)}} &= \int_{\frac{a}{4}}^{\frac{3a}{4}} dx \int_{0}^{b} dy, \ \mathbf{E^{ap}} \cdot \mathbf{e_{m,n,l}} \\[2mm]
&= f\left(..\right) \cdot \int_{\frac{a}{4}}^{\frac{3a}{4}} dx \int_{0}^{b} dy, \ sin\left(\frac{\pi}{b}y\right) sin\left(\frac{n\pi}{b}y\right) cos\left(\frac{m\pi}{a}x\right) \\[2mm]
&= f\left(..\right) \cdot \delta\left(n-1\right) \frac{b}{2} \int_{\frac{a}{4}}^{\frac{3a}{4}} dx, \ cos\left(\frac{m\pi}{a}x\right) \\[2mm]
&= f\left(..\right) \cdot \delta\left(n-1\right) \frac{b}{2} a \left[\frac{1}{m\pi} sin\left(\frac{3\pi}{4}m\right) - \frac{1}{m\pi} sin\left(\frac{\pi}{4}m\right)\right] \\[2mm]
&= f\left(..\right) \cdot \delta\left(n-1\right) \frac{b}{2} \cdot a \left[\frac{3}{4} sinc\left(\frac{3\pi}{4}m\right) - \frac{1}{4} sinc\left(\frac{\pi}{4}m\right)\right]
\end{aligned}
$$

In summary the mode-voltages for each region are,

$$
\boxed{\mathbf{V^{I}}_{\mathbf{(m,n,l)}} = -f\left(..\right) \cdot \delta\left(n\right) b \cdot \frac{a}{2\pi} \frac{4}{\left(m^2 - 4\right)} \left(sin\left[\frac{3\pi}{4}m\right] + sin\left[\frac{\pi}{4}m\right]\right)}
$$

$$
\boxed{\mathbf{V^{II}}_{\mathbf{(m,n,l)}} = f\left(..\right) \cdot \delta\left(n-1\right) \frac{b}{2} \cdot a \left[\frac{3}{4} sinc\left(\frac{3\pi}{4}m\right) - \frac{1}{4} sinc\left(\frac{\pi}{4}m\right)\right]}
$$

E. Applications of Variational Techniques

E.1. Modeling an Open Rectangular Waveguide, with the 'Complexification and Exatrapolation' Technique

E.1.1. Background

Open waveguides have been studied a lot. A review of the various mathematical and numerical methods which have been succesfully used are given very thouroghly by Gardinal in [**?**]. Most notably, variational solutions have been cacluated by M. H. Cohen et. al.[1], and exact solutions by correlation matrix are given by Macphie [**?**].

Recently, Shen and Macphie have presented a technique which provides accurate results and is simpler (in some sense), and computationally more efficient than full-wave solutions. The technique, which was introduced by W. D. Murphy et. al.[2], is appropriatly named '*complexification and extrapolation'*.

Why am I using this technique, when exact methods exist in a variety of different forms? Because I had at my disposal the variational solution to a homogeneously filled waveguide enlargment discontinuity. With this, the additional effort to implement Macphie's '*complexification and extrapolation'* technique was very little. Furthermore, I have not aware of this method being used with the variational technique, which probably is faster than the correlation-matrix method, and may provide other benefits I am unaware of at the moment.

---

[1]The aperture admittance of a rectangular waveguide radiating into half-space, THis is not available on the internet as far as i could find.

[2]Acceleration methods for the iterative solution of electromagnetic scattering problems, I couldnt find this so i just read the abstract.

234

E.1.2. Introduction

The concept of '*complexification and extrapolation*' is explained nicely in [**?**], but will be repeated here for the reader. The technique is as follows, approximate a waveguide radiating into free-space by a waveguide radiating into a larger waveguide. For an accurate solution, a very large waveguide must be used, and the number of modes needed to provide convergence is very large, and thus computationally expensive (this actually crashed my python with Memory Error). The trick is then to fill the larger waveguide with a lossy dielectric. This makes which convergence possible with a reasonable number of modes. This simulation is run for many different values of dielectric loss, and the lossless case is then calculated by extrapolation.

Below is a figure illustrating the waveguide junction used in this approximation, both in 3D and in cross-section. Also shown is the dominant mode, which is used as the assumed field in the variational calculation[3].
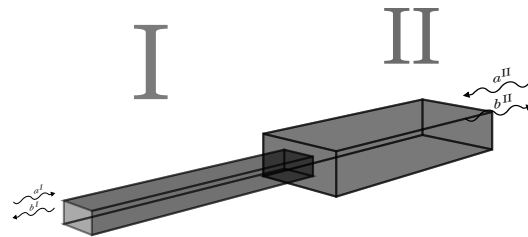
Figure E.1.: 3D illustration of Junction
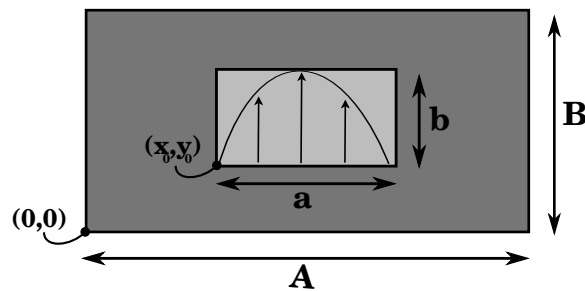
Figure E.2.: Cross section of junction

---

[3]See 'Calculation of the Junction Admittance for a Translation Miss-aligned Rectangular Waveguide Interface' by me for details of the variational calculation.
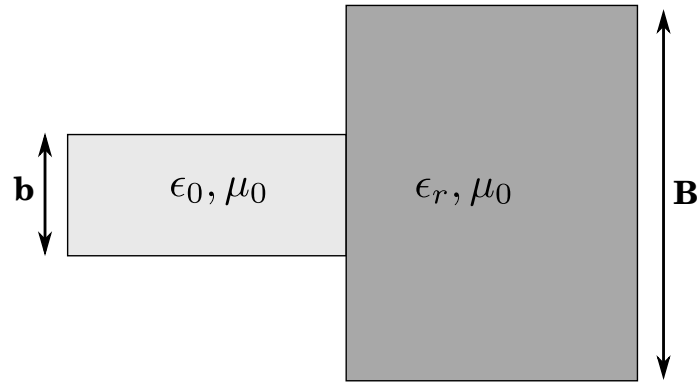
Figure E.3.: Side View of Junction, with lossy Dielectric in Larger Guide

E.1.3. Motivation

To provide an understanding for how the '*complexification and extrapolation*' technique may have be developed, we will examine my failed initial attempts to model the open as a step into a larger guide without loss. Below are variational solutions for the normalized junction addmitance (really the conductance), for the discontinuity illustrated in figure Fig. E.2, with the larger guide width being 10x larger than the smaller guide ($\frac{A}{a} = 10$). Examining this plot we can see many jaggy spikes. These correspond to higher order modes 'kickin on' in the larger guide. The simulations are made for increasing number of modes and one observes that the trace apears to converge, in this case, after about 30 modes are accounted for. (dont worry, a more rigorous measure of convergence is presented later.)
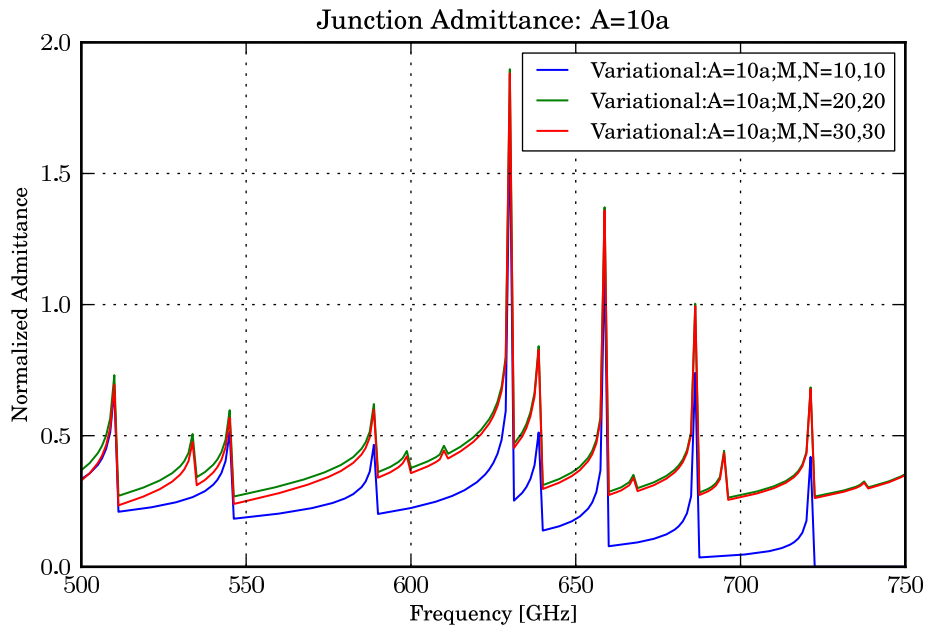
Figure E.4.:

To be confident that the larger waveuide is accurately approximating free-space, this trace should remain stable as the larger guide is increased in size. Below is the same plot as above, except made for larger guide widths equal to 20x*a*, and 100x*a*. What is occuring is more modes are kickin on, and they are smoothing out the admittance.
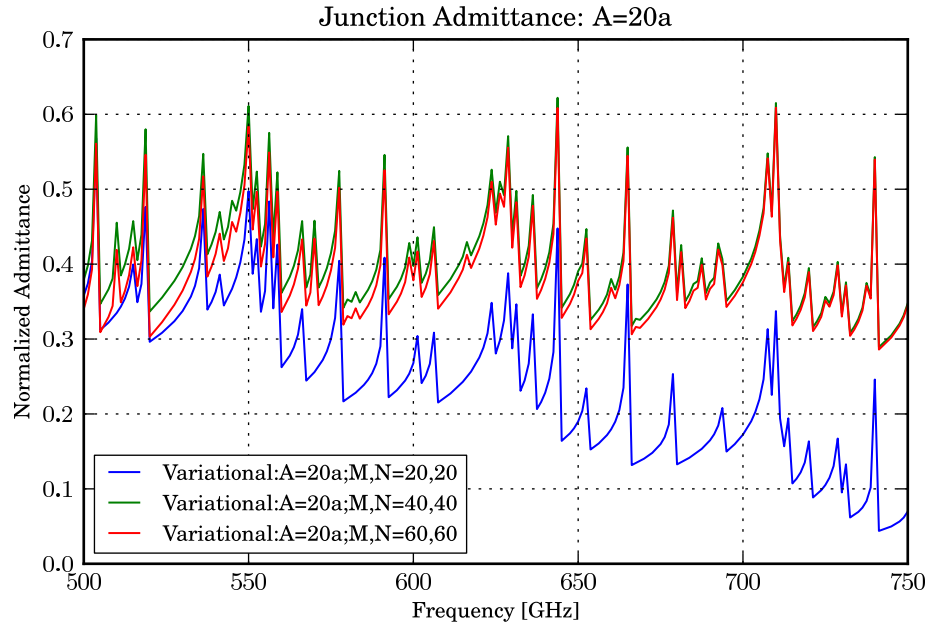
Figure E.5.:
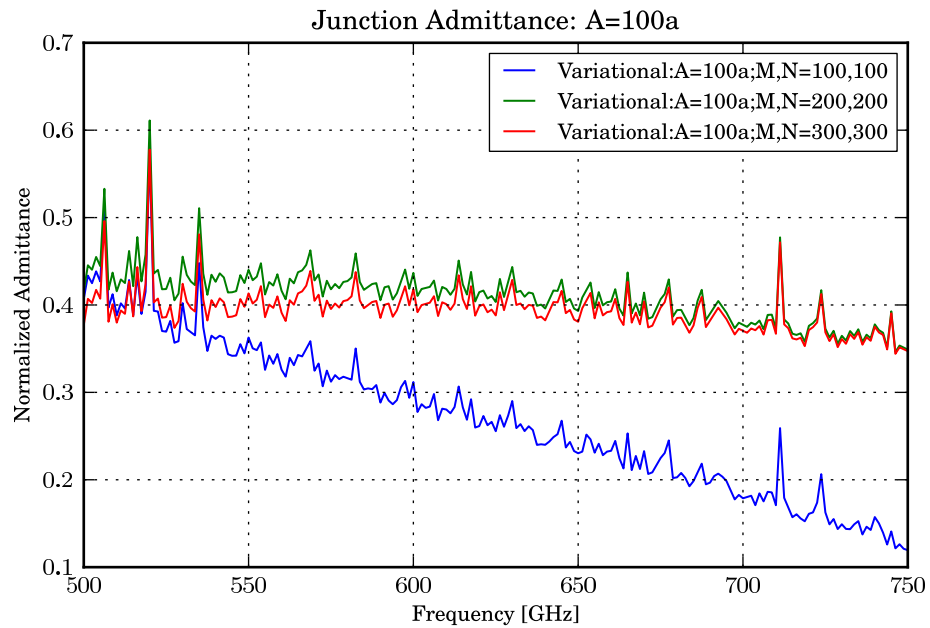
Figure E.6.:

After stairing at these two plots for a few minutes, one realizes; there are two

important quantities here.

1. The size of otuput guide needed for a stable admittance solution, and

2. The number of modes need for a convergent solution, at a given output guide size.

The problem, which is the achillies heel to his approach, is that number of modes needed for a convergent solution grows too fast as the output guide size is increased. This problem can be studied quantitatively once an appropriate measure of convergence is agreed upon. This turns out to be a tricky problem.

E.1.4. Convergence Parameters

The measure of convergence was based off of Bauwen's understanding of HFSS. The convergence parameter was defined as the maximum complex difference between consective admitance values, taken over a band of frequencies.

$$\sigma = max\left(\left|Y\left(\omega\right)_n - Y\left(\omega\right)_{n-1}\right|;\omega\right)$$

The tricky part is that for this value alone, is meaningless. The rate at which the parameter one is varying from sweep to sweep (accross n) must be given to provide meaning to the convergence parameter, σ. To the best of my knowledge, these two values must be chosen huersitically.

To further compound the complexity, there are two convergence parameters we have to deal with, defined as above but for different variables for n.

1. σ: The convergence of the admittance with respect to the number of modes used, at a given scaling factor ($n = \#modes$).

2. ρ: the convergence of the admittance with respect to the 'scaling factor' of the output guide width $(n = \frac{A}{a})$.

Below is a plot of the number of modes needed for a convergence at 625GHz, vs the scaling factor of the output waveguide width $(\frac{A}{a})$ for a variety of dielectric loss values in the larger guide. The convergence parameters in this case where $\sigma = 1e - 3$, and $\rho = 1e - 3$.

This figure illustrates the problem, the number of modes required for convergence vs scaling factor is a linear relationship. Our computation order increases roughly like the number of modes squared. (actually 4x number of modes squared, because I take same number of modes in each guide, and there are 'TE' and 'TM' modes). This combined with a plot of the admittance with respect to the scaling factor for a lossless, is a non-rigorous proof that this procedure is doomed.
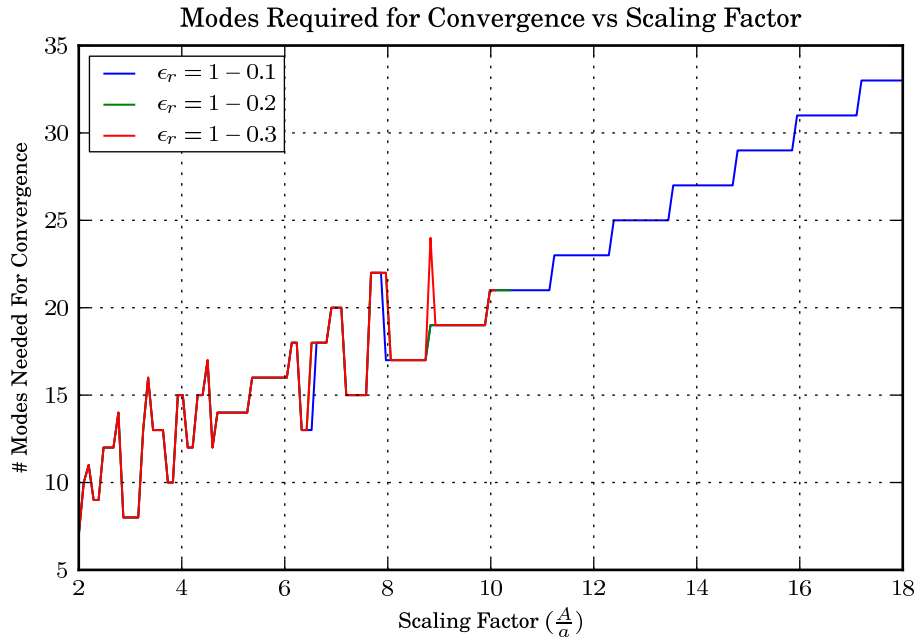
Figure E.7.:

If loss is introduced into the larger waveguide, however, the convergence of the

admittance with respect to the scaling factor is much improved. Below are plots of real and imaginary parts of the admittance for 3 dielectric loss values calculated at 625GHz. Similar plots are given in [**?**]. As one can see, the larger the loss the quicker the convergence.
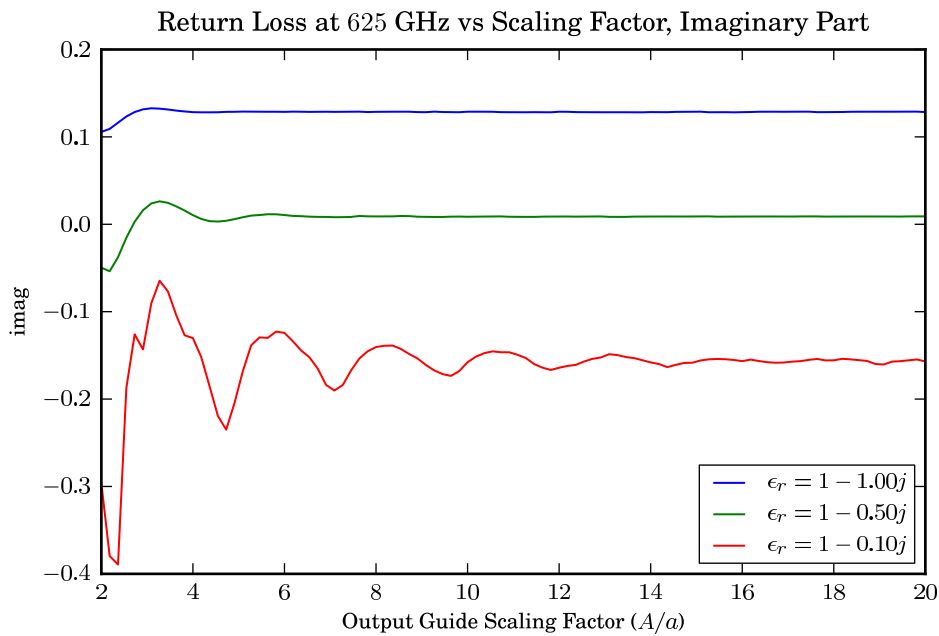
Figure E.8.:
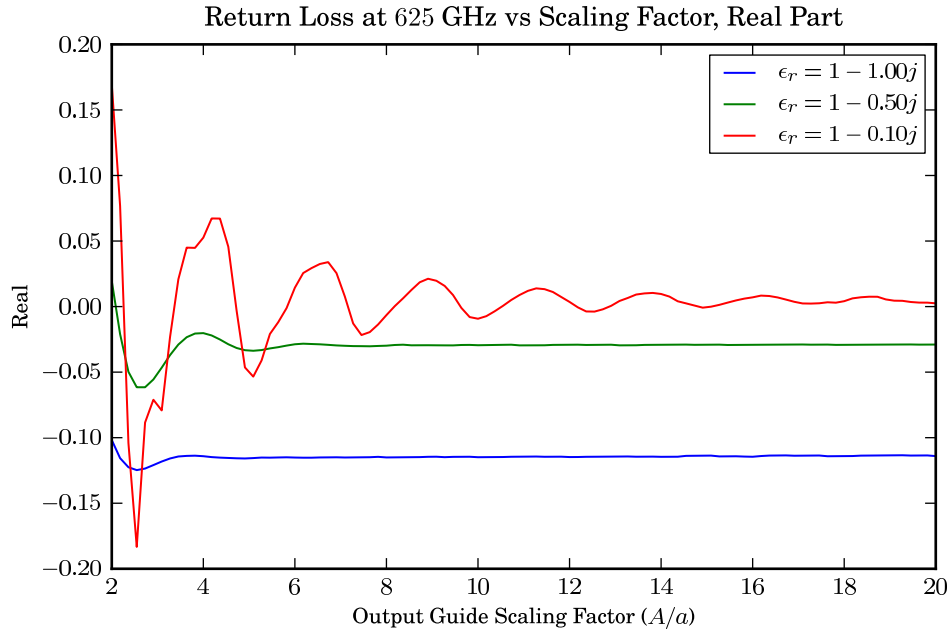
Figure E.9.:

The above plots are components of the complex admittance plot below, which is a parametric function dependent on scaling factor.
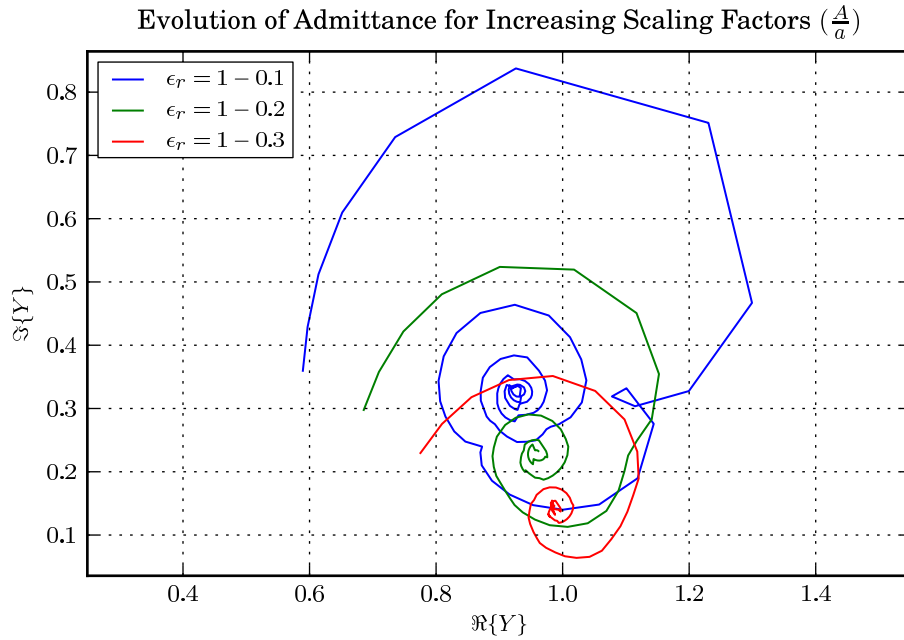
Figure E.10.: Junction Admittance, as Scaling Factor Increases

E.1.5. Exatrapolation

The idea now is to get three converged results (converged with respect to both σ and ρ) and use a lagrange polynomial extrapolation to calculate the admittance value in the lossless case.

A useful tip from [**?**] about the polynomial extrapolation,

As pointed out in [..], parabolic extrapolation is second-order accurate provided the function .. is sufficiently smooth.

Meaning if the dielectric losses chosen are linearly spaced from 0 onward, the error in the lossless case of good to the second order. For the extrapolation I used the same dielectric loss values as in [**?**]. Below are plots of the Conductance and Susceptance, as above, but for the dielectric loss values I used in the interpolation. For the interpolation calculation used the a function from scipy (scipy.interpolate.lagrange)
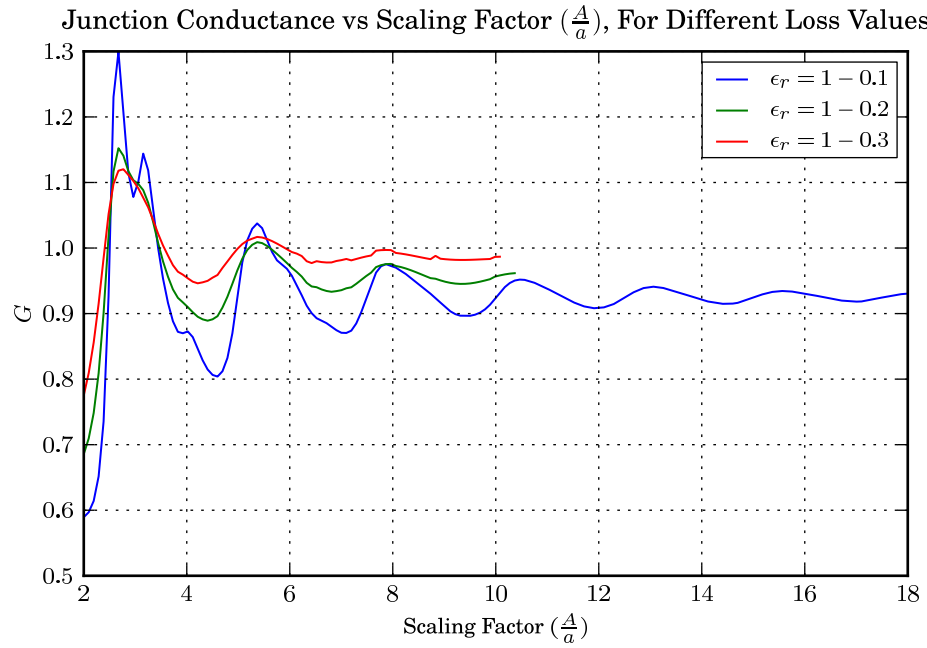
Figure E.11.:
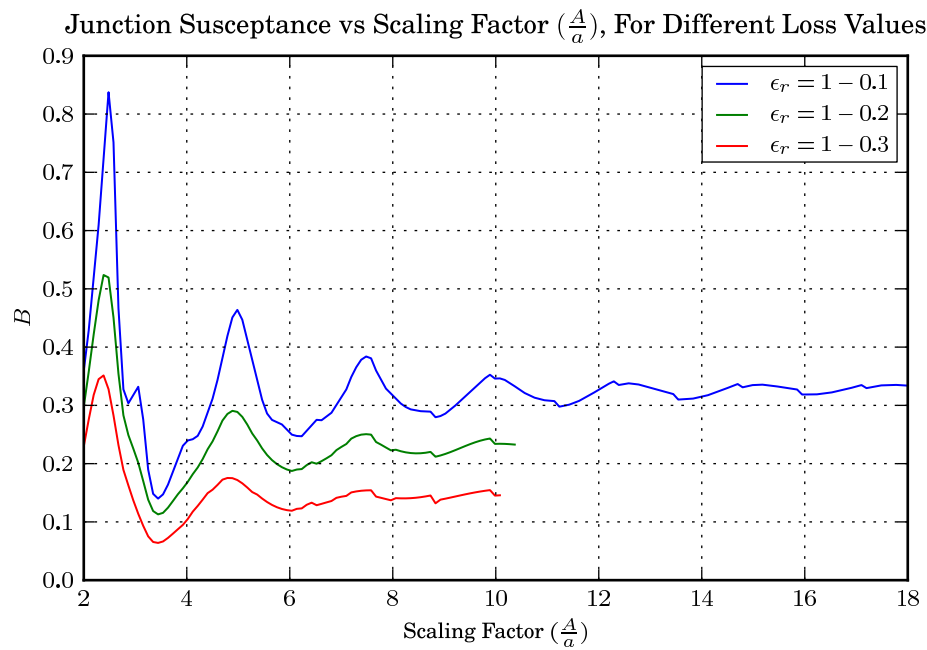
Figure E.12.:

Below is a plot illustrating the extrapolating polynomial at 625GHz.

Figure E.13.: Extrapolating Polynomial

E.1.6. Comparison With HFSS

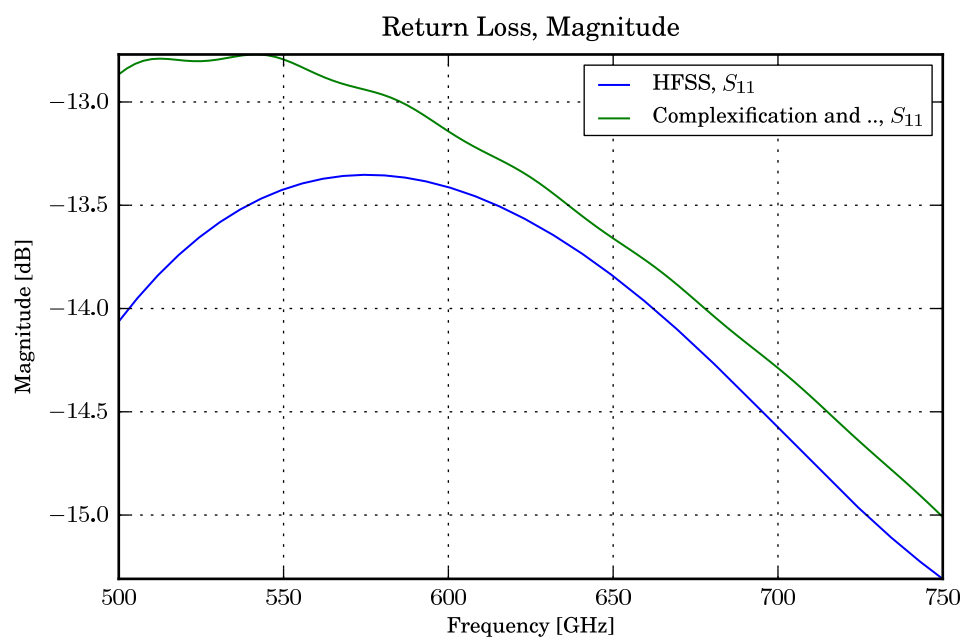Below are the simulation results compared with HFSS.

Figure E.14.:

Junction Susceptance
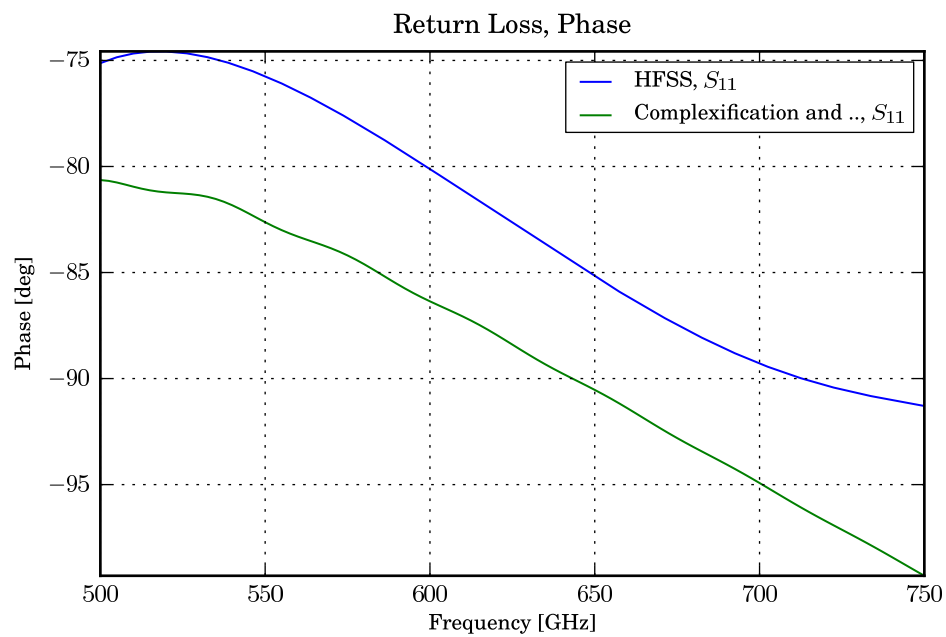

Figure E.15.:

Return Loss, Magnitude
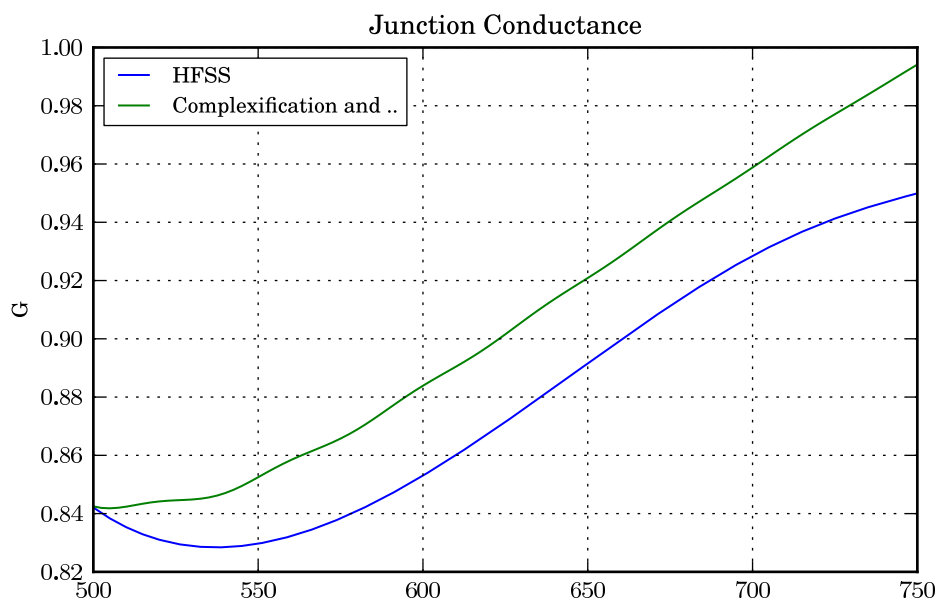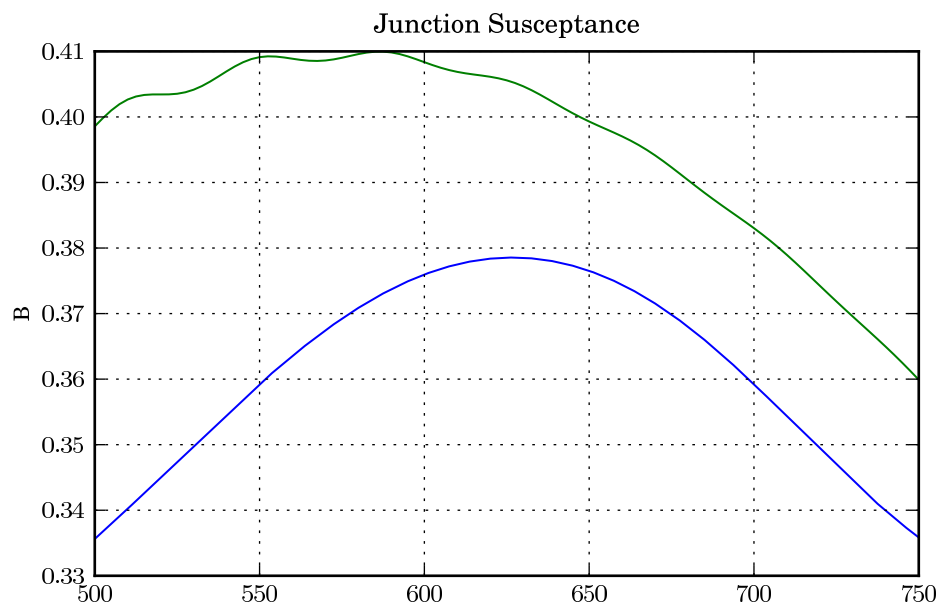
Figure E.16.:

Figure E.17.:

E.2. Measurements of the Radiating Open Standard at WR-10

E.2.1. Introduction

The usage of the rectangular waveguide radiating open (RO) as a calibration standard has many desirable qualities, such as

- Resistant to flange alignment verfication

- Its free

- well approximaited by simple models at higher frequencies

- Provides common impedance standard.

Recently, efforts led by Dylan Williams to produce an accurate analytical solution for a RO with inclusion of flange effects to be used for calibration have been made[4]. In contribution to these efforts, measurements verifying the solution at many waveguide bands are being made.

This paper presents measurement results for the WR-10 waveguide band made using an HP-8510 with HP-XXXX extension modules. These measurements include comparisons between a radiating open from a standard UG-387 flange and a radiating open from an approximation to a infinite flange.

E.2.2. Setup

Below is a picture illustrating the general measurement setup. Specifically this was taken during the measurement of the radiating open standard. Close-up pictures illustrating each flange are seen in Figures 2 and 3.

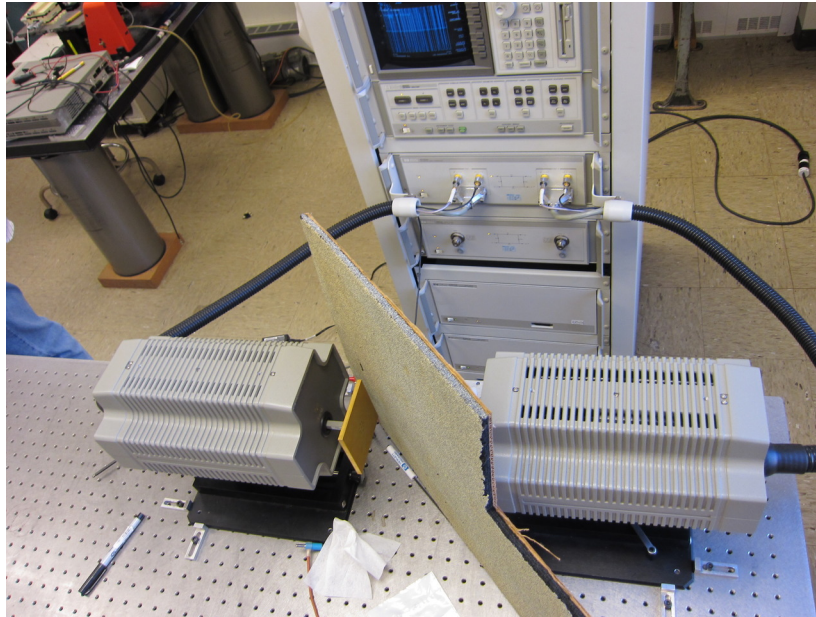[4]Reflection Coefficient of Radiating Rectangular-Waveguide Test Ports
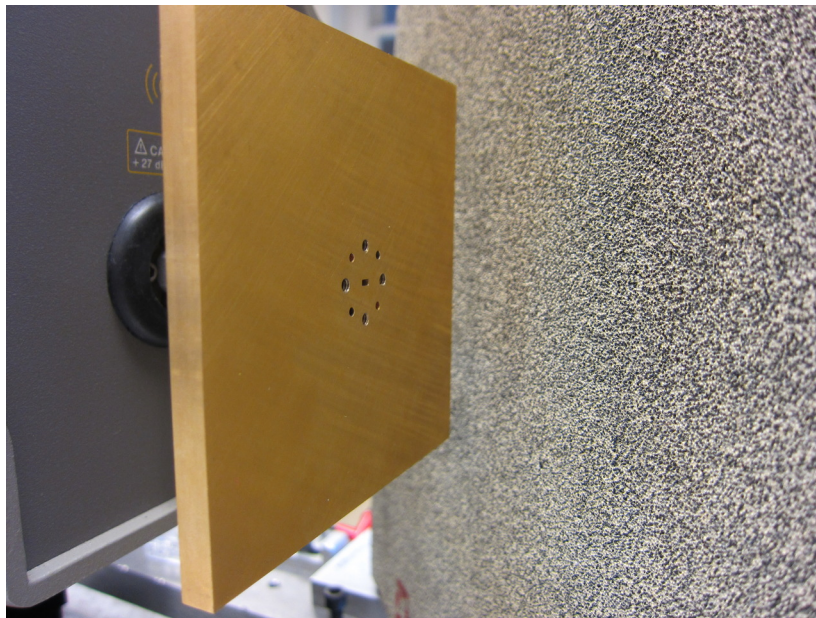
Figure E.18.: Picture of Setup

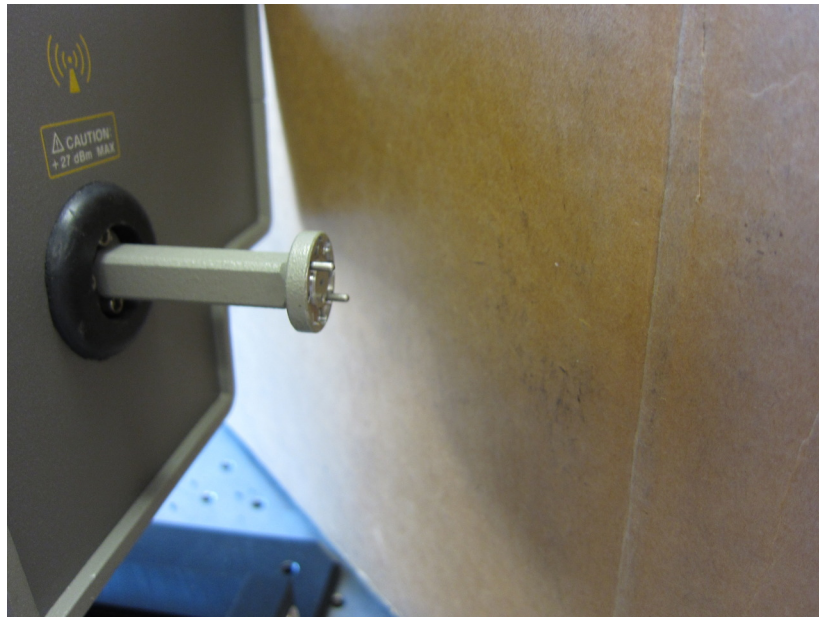

Figure E.19.: Close up of Approximation to Infinite Flange

Figure E.20.: Standard Flange

E.2.3. Initial Results[5]

To make the following explanation clear, my naming convention is as follows:

Transmissive standards are described by a single word or phrase, while reflective standards are described by a comma seperated word/phrase, each describing the one-port standard on a single port. For example; *(short, match)* is interpreted as meaning 'short on port 1, and a match on port 2'.

The standards measured are:

- thru

- line

- short, delay short

- delay, short short

---

[5]These measurements where made on 01/21/2011.

- match, short

- short, match

The line and delay short standard are realized using a quarter wavelength shim at band-center, $(d = \frac{\lambda_{bc}}{4})$. Calibrations were calculated using a generic least squares algorithm that neglects leakage. One calibration set , refered to as 'without Match', set consisted of

- thru

- line

- short, delay short

- delay, short short

The other calibration set used, as 'with Match', consisted of

- thru

- line

- match, short

- short, match

The results of the infinite flange as well as the standard flange are shown below for each calibration set.
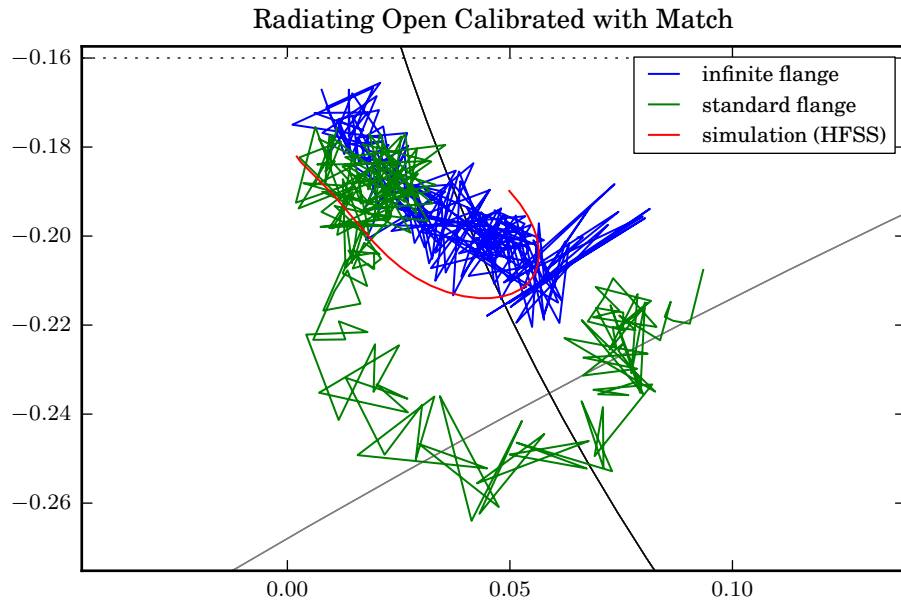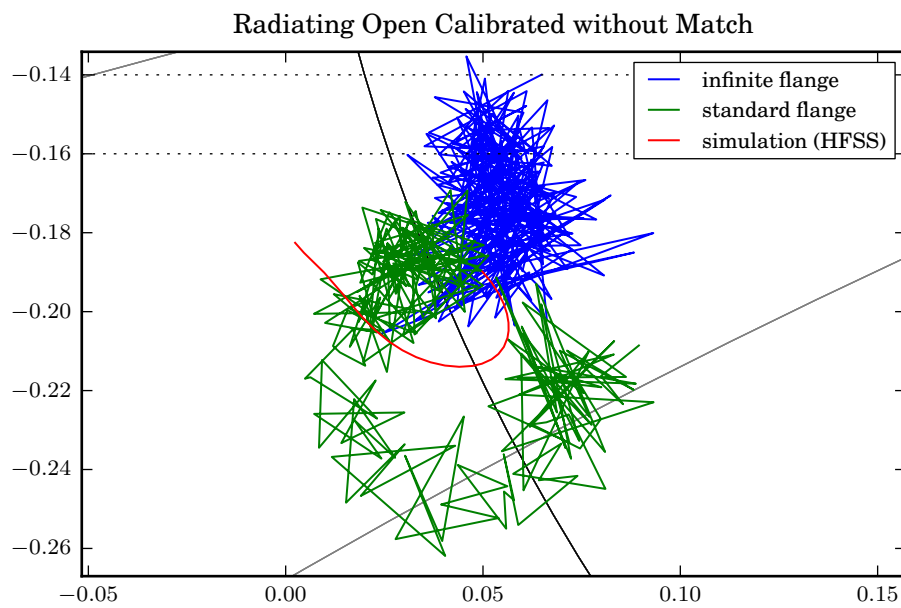
Figure E.21.:

Figure E.22.:

The results indicate that the simple HFSS model predicts results that are on the

order of the measurement noise, and variation between calibration sets. Thus, it is hard to conclusively state the effect that the pins and finite flange have on the reflection coefficient for the open. However, loosly speaking, the standard flange seems to produce a slightly larger and more varied reflection coefficient.

Of the two calibrations and both flanges, the calibration 'with match' shows that the infitinite flange follows the simulated response most closely.

E.2.3.1. Verification plots

These plots are calibrated measurements of standards using in the calibration set. The de-trended phase plots are de-trended from the ideal definitions, which are calculated analytically.
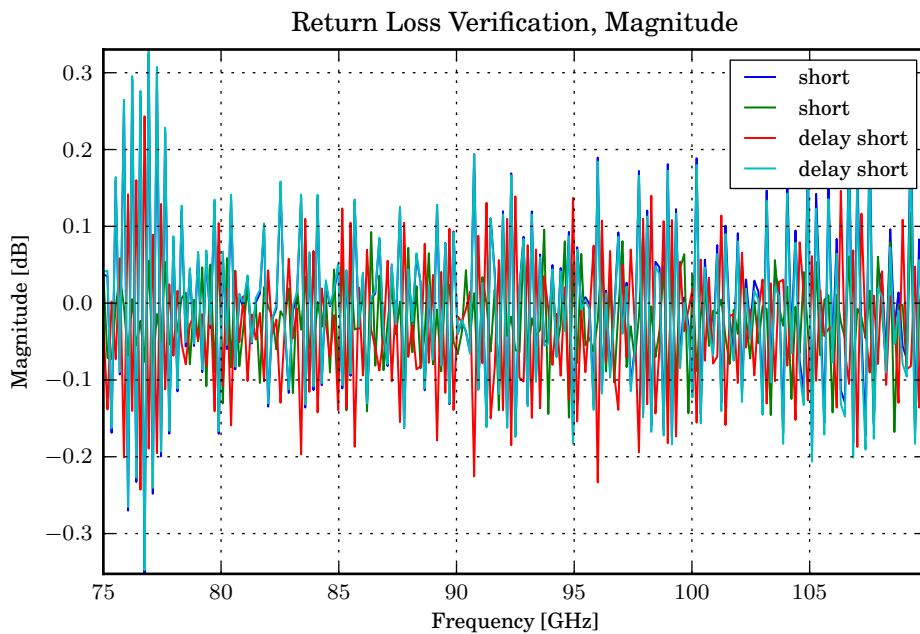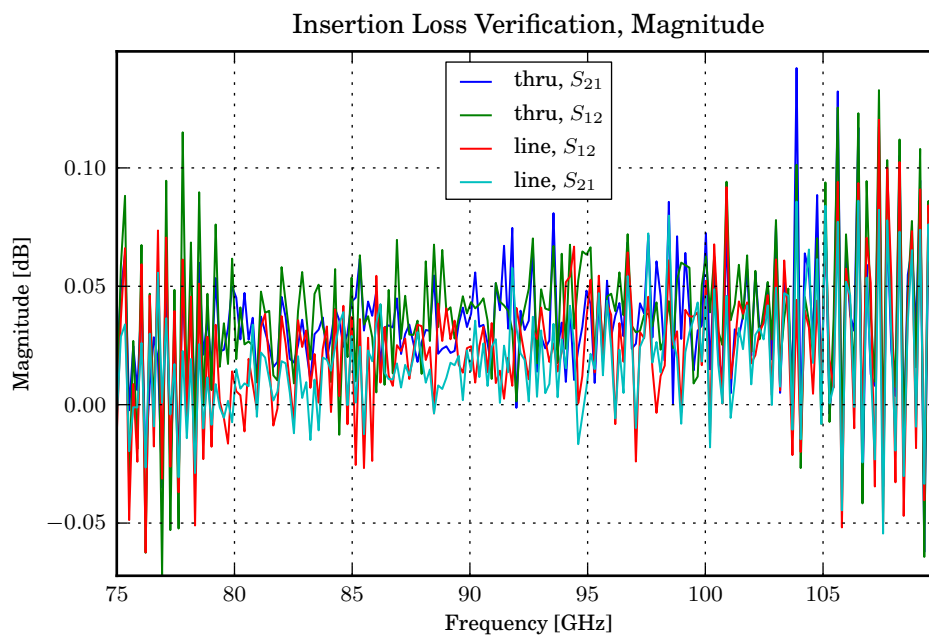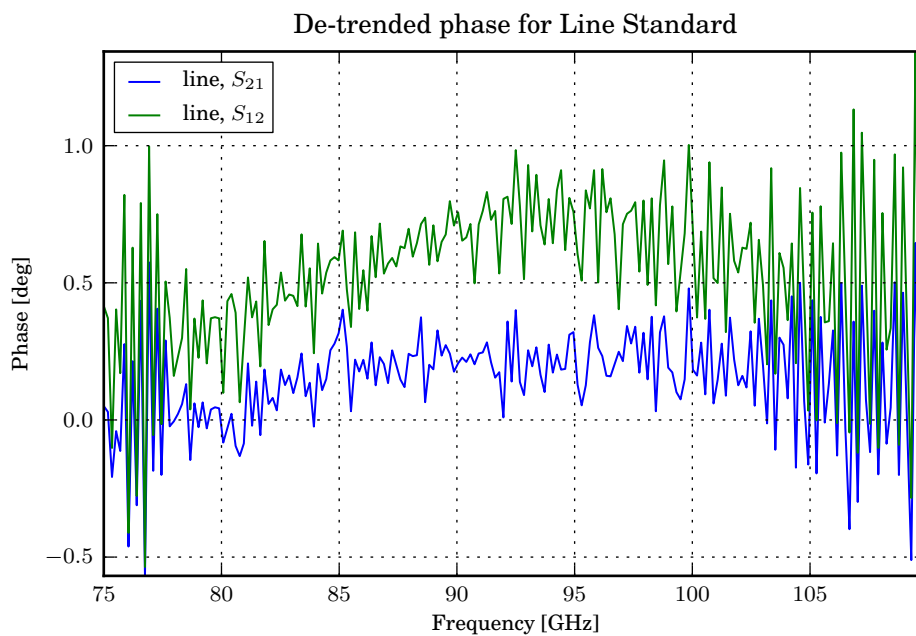

Figure E.23.:
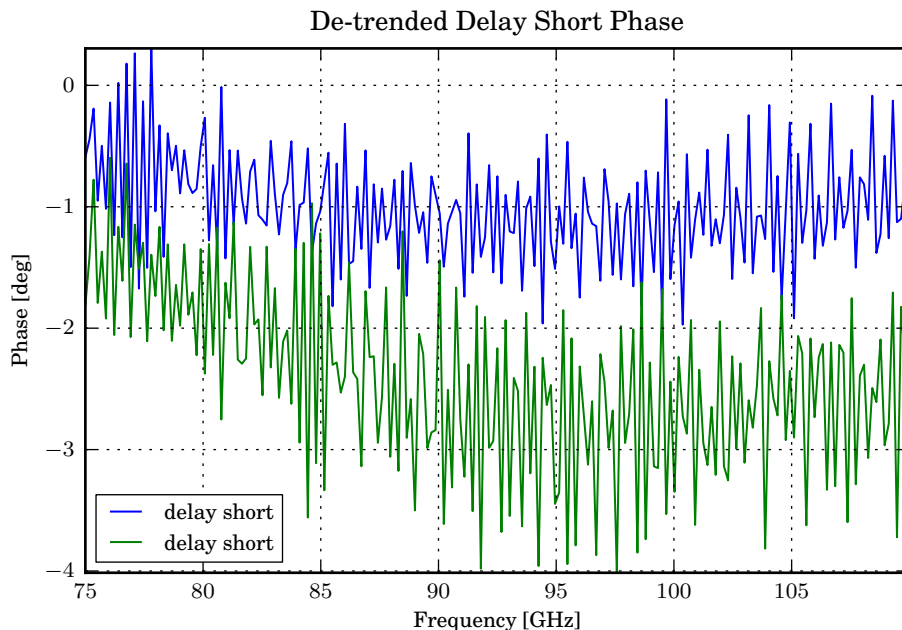
Figure E.24.:

Figure E.25.:

Figure E.26.:

E.2.4. Improved Results [6]

E.2.4.1. Switch-terms

After a helpful discussion with Dylan, the 'noise' in the last section was discovered to be largely due effects of the 'switch terms'. See [24] for full details on switch terms. In short, the switch terms are extra error terms which represent the fact that the VNA does not fit the 'error-box' model perfectly. This is because for VNA's with one source, the non-sourced port must be terminated with an internal load. The load is not ideal, and is different for each port (actually this is due to an electronic switch). However, the response of the imperfect load can be measured and its effects removed. Measurement ability for switch terms in the HP8510, as well as general suport for 'un-terminating' switch terms measurements during calibration was implemented in

---

[6]These measurements where made on 01/23/2011.

mwavepy.

E.2.4.2. Effects of Using Switch-terms

The effects of taking into account the switch-terms where measured. This illustrated in the plots below. Each graph contains three traces:

1. two-port: two-port calibration as described above.

2. one-port: a one-port calibration at each port using: short, delay-short, and a load.

3. two-port with switch-terms: two-port calibration after switch-terms have been un-terminated from the raw measurements

The results show the effect of using switch-terms is significant. The fact that the standard two-port error seems to be bound by the one-port calibration is interesting and I can not think of an explanation.

The verification standard used was a Rotated Delay-Short (RDS), which should have unity reflection. The phase of this standard was predicted using both HFSS models, and a variational solution.
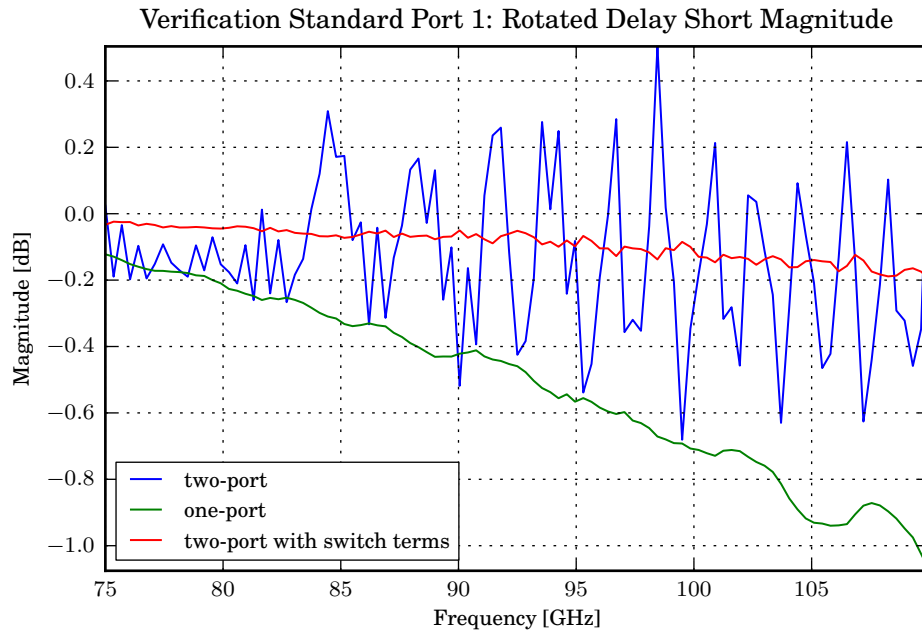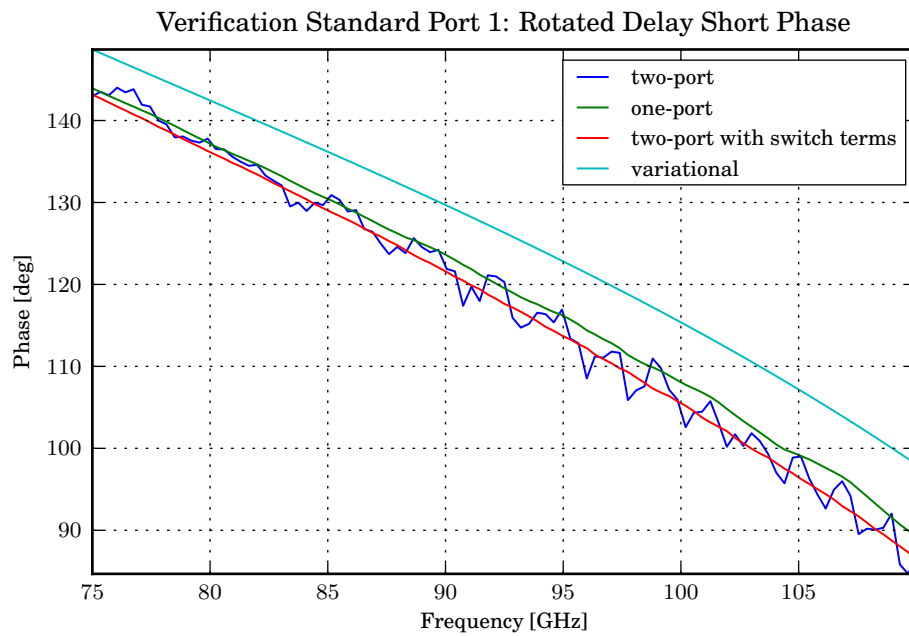
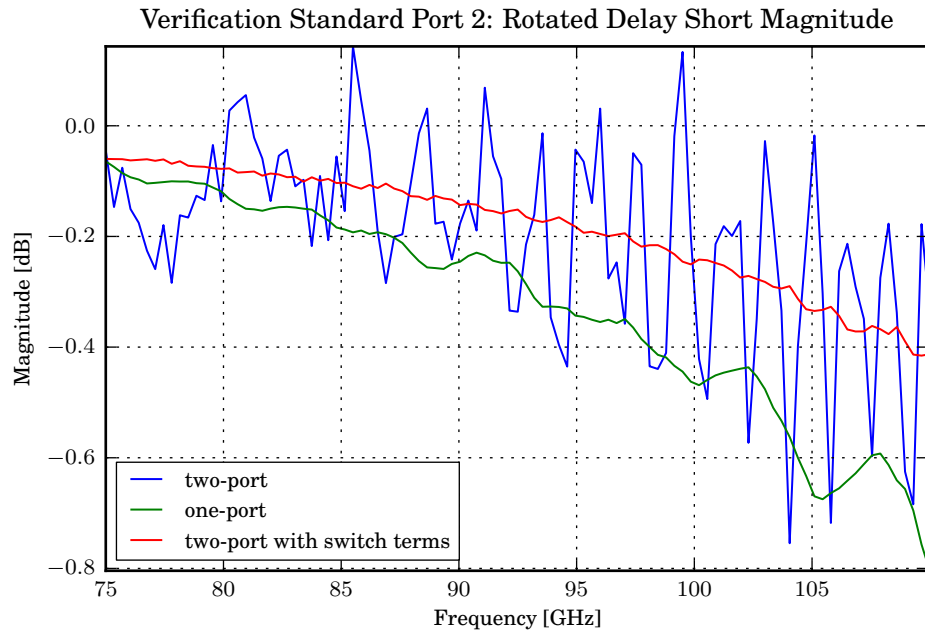**Figure E.27.:**

Figure E.28.:

Verification Standard Port 2: Rotated Delay Short Magnitude

Figure E.29.:

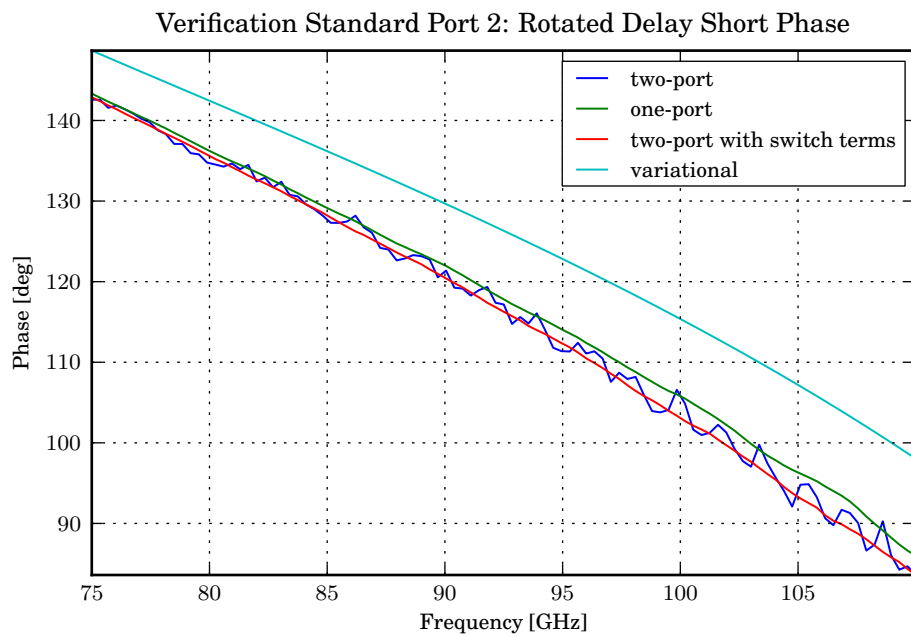Verification Standard Port 2: Rotated Delay Short Phase

Figure E.30.:

E.2.4.3. Results

The reflection coefficient and its polar components for the RO are shown below. The labels given in the legend are explained below.

1. inf flange: large plate which approximates an infinite flange, as shown in figure Fig. E.19.

2. std flange: the standard UG-387, as shown in figure Fig. E.20.

3. HFSS: simple HFSS simulation by alex

4. From Dylan: complex HFSS simulation from dylan

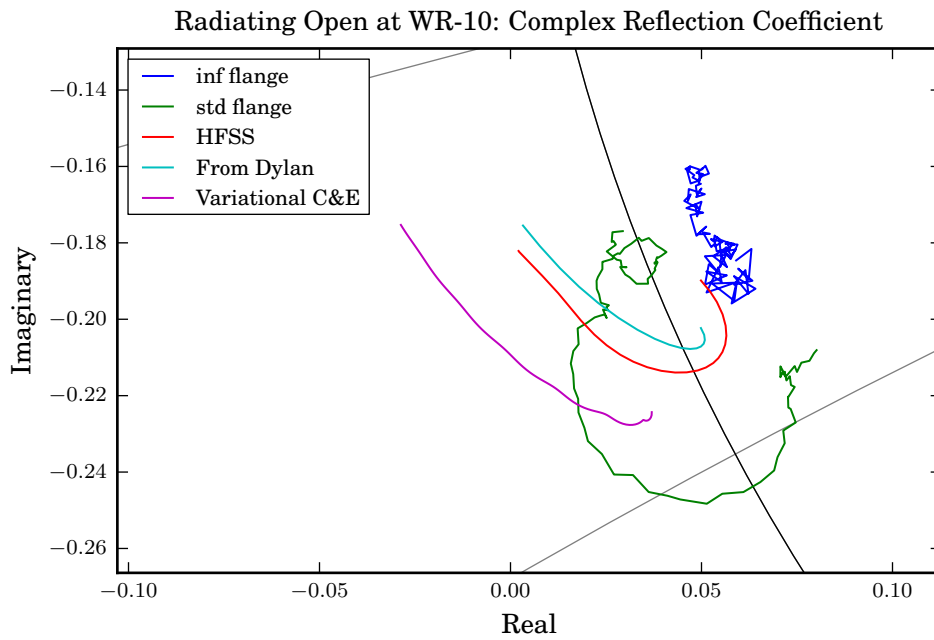5. Variational C&E: a Variational Complexification and Extrapolation technique, as described in [7]

Figure E.31.:

---

[7]Modeling an Open Rectangular Waveguide, with the 'Complexication and Exatrapolation' Technique. Alex Arsenovic 10/22/2010
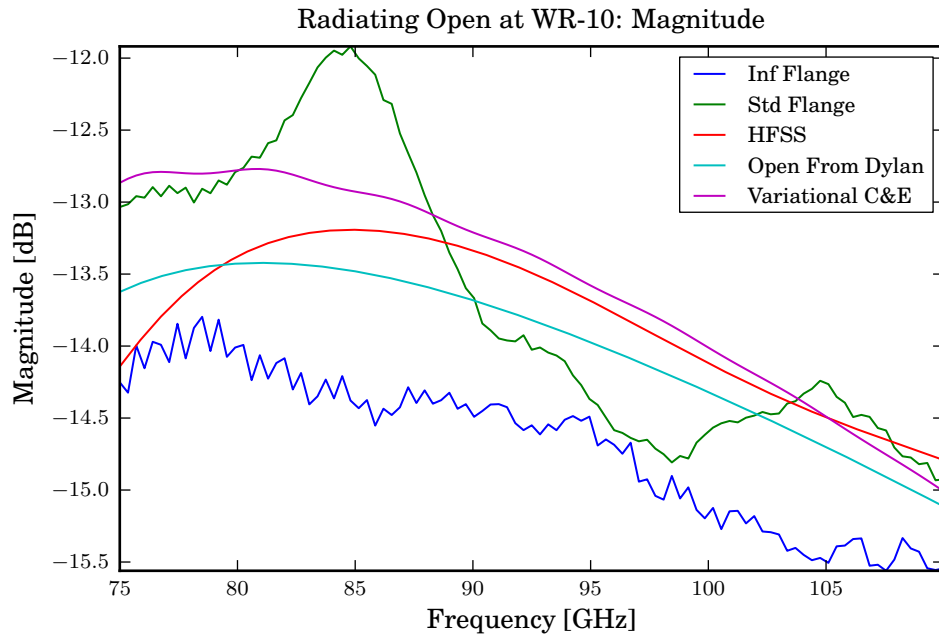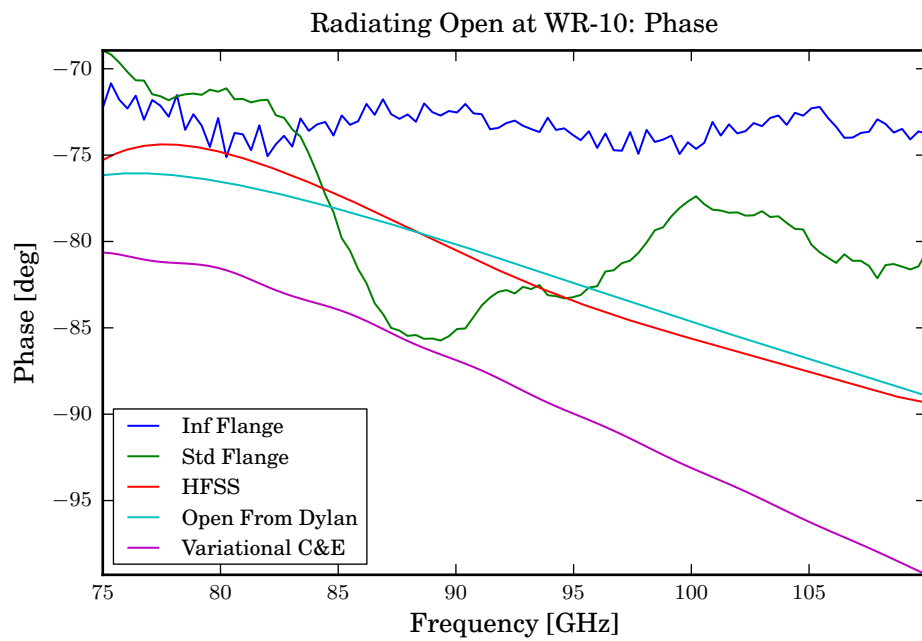
Figure E.32.:

Figure E.33.:

E.2.4.4. Interpretation

Firstly, the measurements can interpreted simply as a verification of an analytical model. From this persepective, it is appropriate to compare the inf flange to the HFSS simulations. This can be done visually through inspecting the plots in the previous section. Alternatively, it can be quantified in terms of a metric that is more relevant to the problem of calibration; the complex distance between the two responses. This is shown below in figure Fig. E.34 (note that the inf flange is re-termed 'Plate')
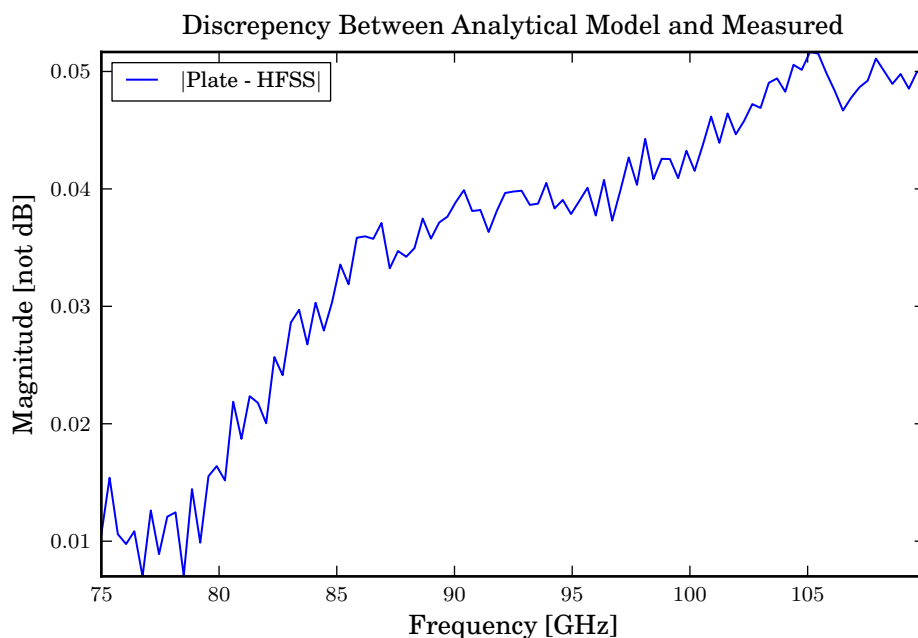

Figure E.34.:

While the previous figure is important for validating the analytical model, accurately modeling a large plate is not useful for practical calibration applications. Instead, we need a model for the actual UG-387 flange. Because the flange is geometrically complex, it is desirable to model only the components that are most significant to its electrical response.

An appropriate question to ask is; how different is the flange and infinite plate? Shown in figure Fig. E.35 is the magnitude of the complex difference between measurements of the flange and the infinite plate.
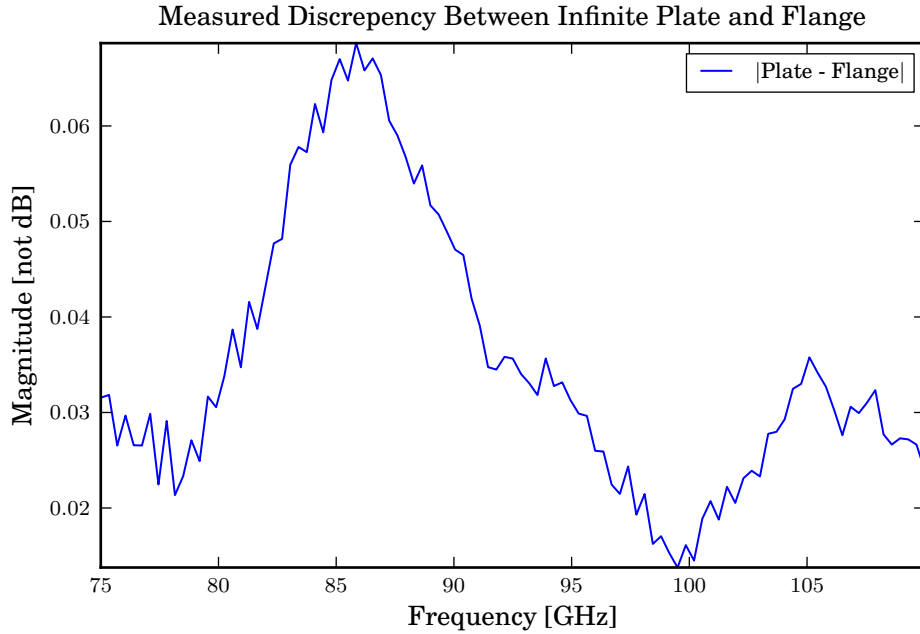

Measured Discrepency Between Infinite Plate and Flange

Figure E.35.:

Comparing figure Fig. E.35 with figure Fig. E.34, we can conclude that the magnitude of the difference between our model and measurement of the infinite plate, is comparable to the difference between the infinite plate and the flange. In other words, the accuracy of our model is comparable to the effects of the flange.

This brings us to a somewhat strange but practical comparison. Even though our analytical model is for an infinite plate, how far is it from the measurement of the actual flange? This difference is shown in figure Fig. E.36. Suprisingly, figure Fig. E.36 indicates that theorectical response of the infinite flange is closer to response of the standard flange. This may be attributed to two causes; either the infinite plate is far from what we are modeling in HFSS, or the measurements are not accurate.

Whatever the reason, it is hard to put theoretical efforts towards improving model of the actual flange, when the model for the geometrically simpler inf plate is as inaccurate as the two measurements are different.
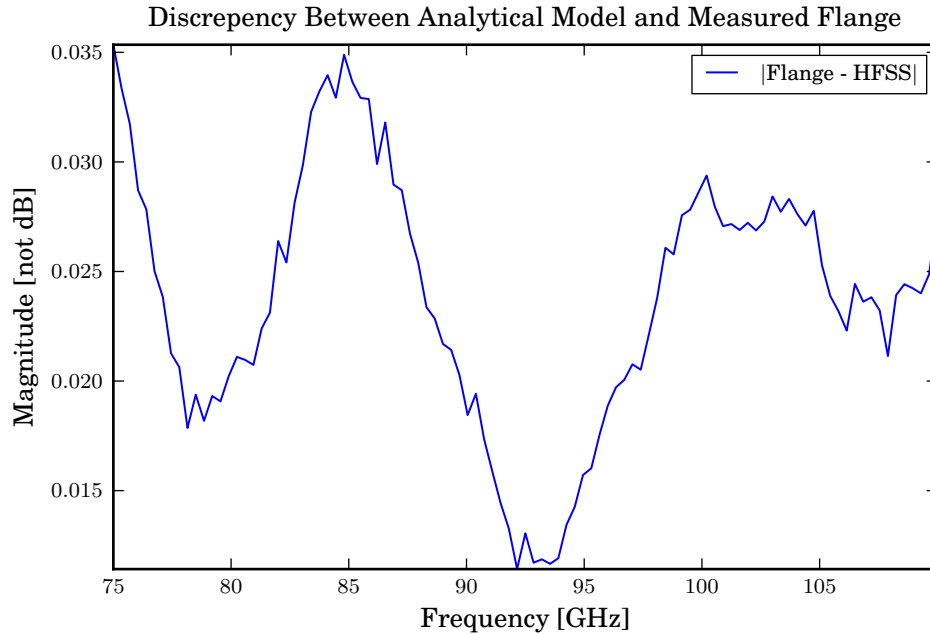

Figure E.36.:

E.2.4.5. 'Good Enough'

Although improving the model for the RO may be justified in and of itself, a well defined goal in accuracy is important to define before the standard is used in practice. One of the main benefits of the RO as a load standard is the alieviation of flange misalignment. Therefore, a reasonable goal is to achieve an agreement between theory and measurments that is bettter than the error introduced by the flange for a load standard.

Shown in figure Fig. E.37 is a comparison of the current model analytical-measurement descrepency against two simulations of misalignment on a perfect match standard.

1. Misaligned Match 1: perfect match with missalignment of $\delta_a, \delta_b = \frac{a}{10}$

2. Misaligned Match 2: perfect match with missalignment of $\delta_a, \delta_b = \frac{a}{30}$
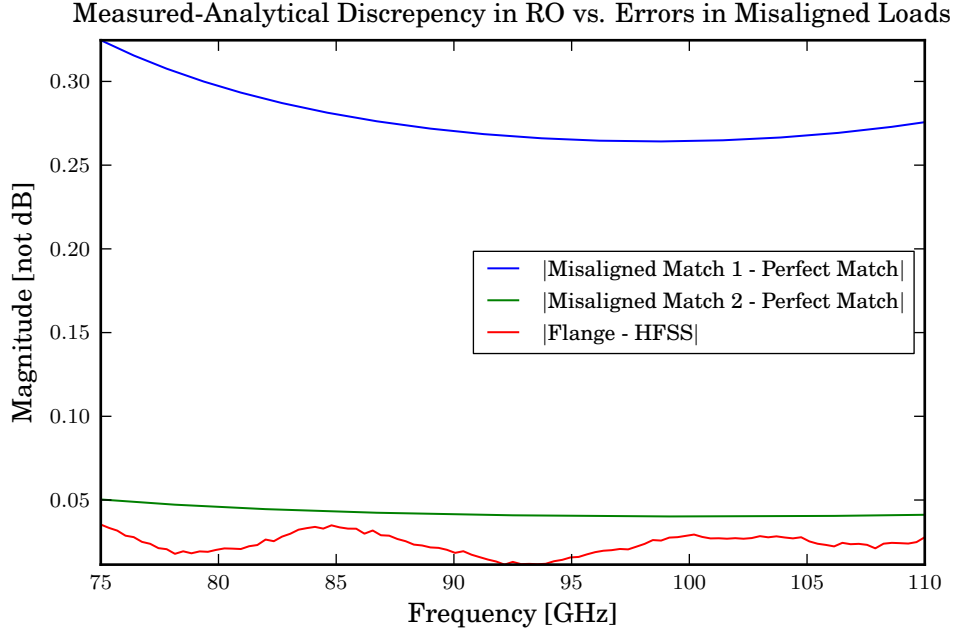
Figure E.37.:

From this plot, it can be concluded the effects of the simulated waveguide missalignments are more significant than the inacuracy of the RO model. This conclusion relies on the assumption that our measurements faithfully represent the true response of the RO.

In other words, our model for the RO is in error, but this error is less than the error introduced by a waveguide missalignment of $\frac{a}{30}$.

Complex plots of the different load standards are shown below.

Figure E.38.:

E.3. Self-Calibration of Misaligned Rectangular Waveguide

E.3.1. Summary

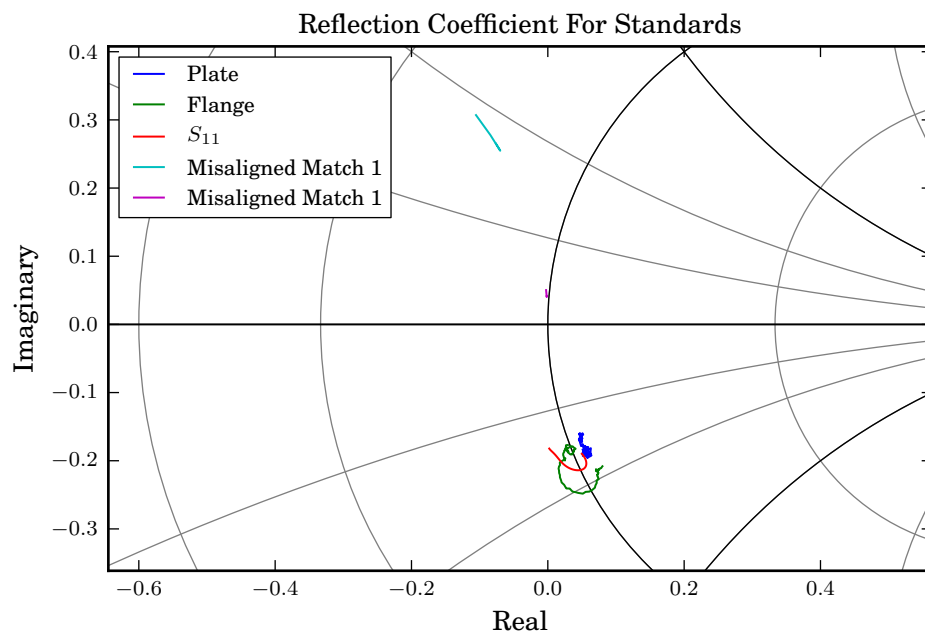A virtual experiment, has been constructed to test the performance of a self-calibration routine in presence of flange-induced translation missaligment errors. The results for the initail test are shwon in detail. Following this, a monte-carlo simulation was conducted to determine the algorithms average performance, given a realistic model for the offset.

The results show that the procedure does work. How well?

- Given offsets of approximaitley $\frac{a}{10}$ in each dimension, the offsets where determined to within about 20% of their true values (within 5% for the monte-carlo analysis).

- The improvement in determined embedding network scattering parameters error, is approxmaitly an order of magnitude.

The results are encouraging, and warrant further investigation, but must be interpreted within the correct scope. This experiment has shown that the procedure works in pressence of noise, given that the the flange miss-alignment is the ONLY source of systematic error. It may also work in presence of other systematic errors, with or without a modified self-calibration scheme, but this cannot be assserted from the following results.

E.3.2. Explaination of Setup

- A ficticous, but reasonable, 2-port error network was used to simulate the systematic effects of the VNA.[8]

- A calibration was performed using the following standards; short, open, match, $\frac{\lambda_{bc}}{4}$ delay short. It was arbitrarily simulated at X-band, but the entire process scales seemlessly to any band.

- The short and open where both assumed to be perfect. The match and delay short where simulated with randomly selected, but known, translation offset of magnitude approximaitly $\frac{a}{10}$ [9]. This value is taken from measurement estimates at WR-1.5[10]

- Standards where cascaded behind the ficticous VNA to simulate measurements. Realistic, additive, white gaussian noise, based off previous measurments of VNA drift was then added to the measurements[11]. Note that this noise has different distrobutions in magnitude and phase components.

- The measurements where then used in a standard least squares calibration and a self-calibration routine with unknown translation miss-alignment[12].

- *Algorithm Configuration:* The self-calibration was provided $\left(\frac{a}{20}, \frac{a}{20}\right)$ as the initial guess for offsets. The function and parameter vector tolerance (ftol,xtol) passed to the minimization algorithm was 1e-5. The variational calculation used to calculate the response of the offset junction, was set to converge the maximum succeeding normalized difference in junction addmittance to $max\left(\Delta_{Y'}\left(n, n-1\right)\right) = 1e-3$.

---

[8]The S-parameters for VNA was taken from a previous calibration.
[9]see my variational expression for waveguide offsets
[10]From Hesler and Dylan
[11]see my VNA noise analysis
[12]see my paper on self-calibration

E.3.3. Results

The true, and self-calibration determined offsets for the delay and match standards are plotted below. The determined offsets are very close to the true values, with an error of about 20% of the magnitude of the offset. This data is interesting two reasons:

1. it provides verification that the algorithm finds the true values, and

2. it provides a way to measure flange missalignment, by using the flanges themselves.
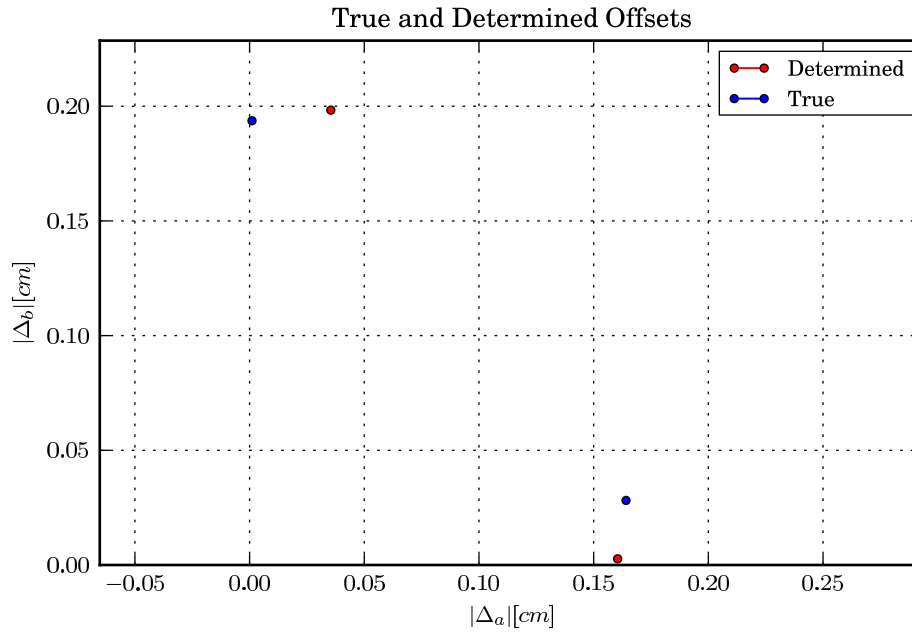

Figure E.39.: True vs to Determined Offsets

Below is a a plot of the magnitude of the errror in determined scattering parameters of the embedding network. The self-calibration routine shows an order of magnitude improvement over the traditional least squares approach.
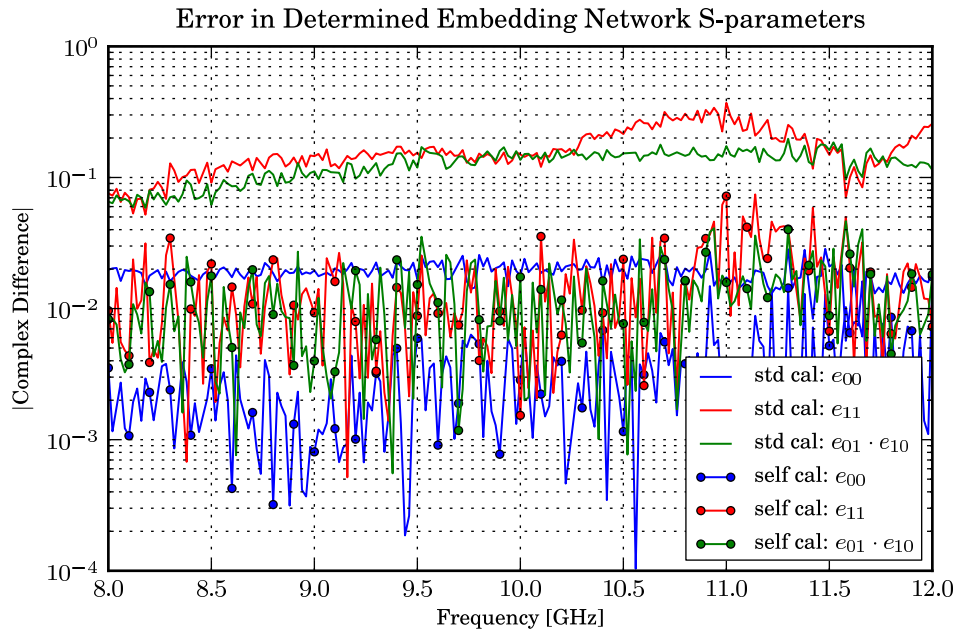
Figure E.40.: Error in Scattering Parameters of Determined Embedding Network

For practical figures of merit, which more closly resembles somthing you could witness in an actual measurement, calibrated measurements of verification standards are given. The following is calibrated measurements of a perfectly aligned $\frac{\lambda_{bc}}{8}$ shim, ad a perfectly aligned match.

Figure E.41.: Calibrated Measurement of $\frac{\lambda_{bc}}{8}$ shim (not used in calibration).

Figure E.42.: Calibrated Measurement of Perfect Match (not used in calibration).

E.3.4. Monte-Carlo Statistical Analysis

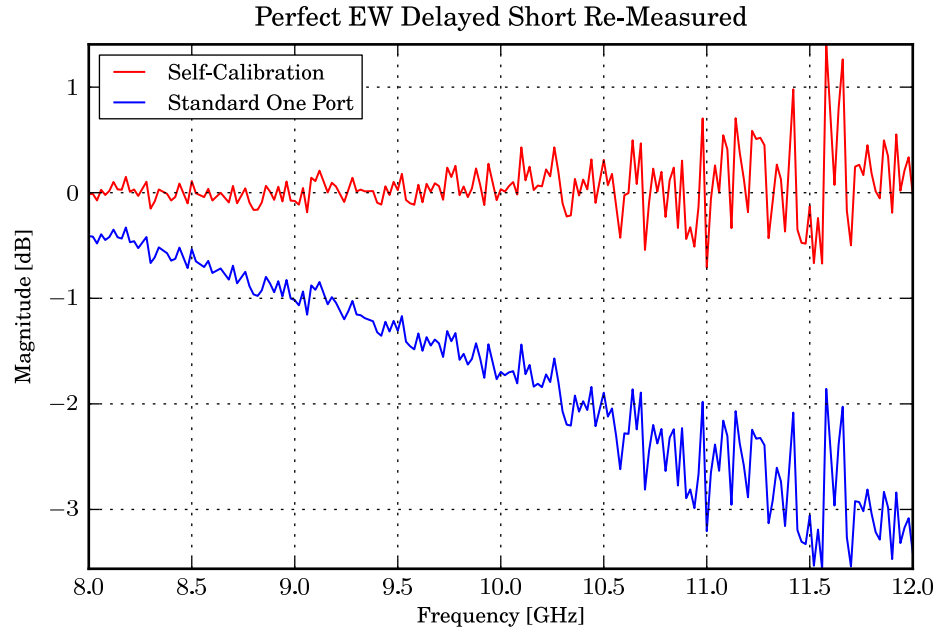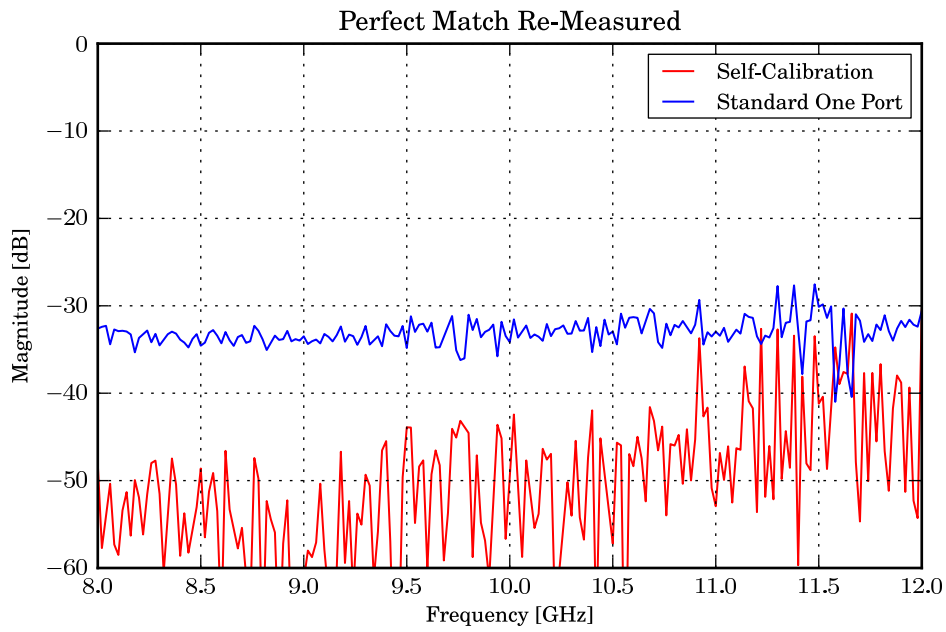A statistical analysis was conducted to ensure that the above performace was not abnormal for the routine. The same calibration as described above was conducted 101 times, which means 201 offset flange interfaces. The flange offsets where selected as random variables from a physically realistic distrobution. The magnitude was sampled from an arc-sine distrobution[13], with largest offset being $\frac{a}{10}$. The phase was selected from a uniform distrobution between $0, 2\pi$. A histogram of the randomly chosen offset magnitudes, is shown in the plot below.

Figure E.43.:

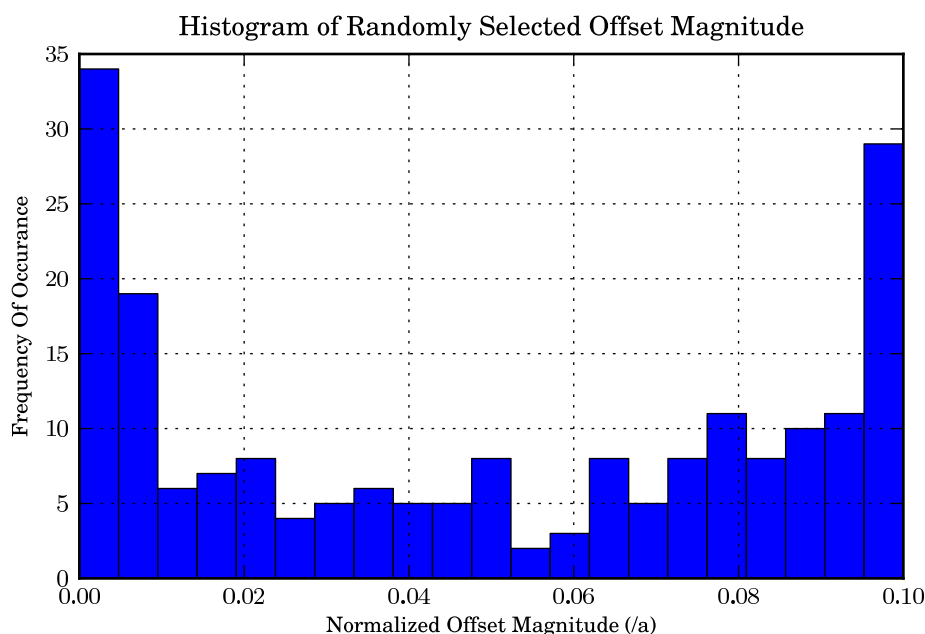The normalized error in determined offset was measured as the euclidean distance between the actual offset and the deftermined offset. A histogram of the normalized error is shown below. From this sample, the normalized error was all below 5%. This is better than originally simulated. Offsets of magnitude $< \frac{a}{100}$ where removed

---

[13]see dylan's 500-750ghz waveguide calibration paper

from this histogram because their normalized error is exagerated due to their small magnitue.
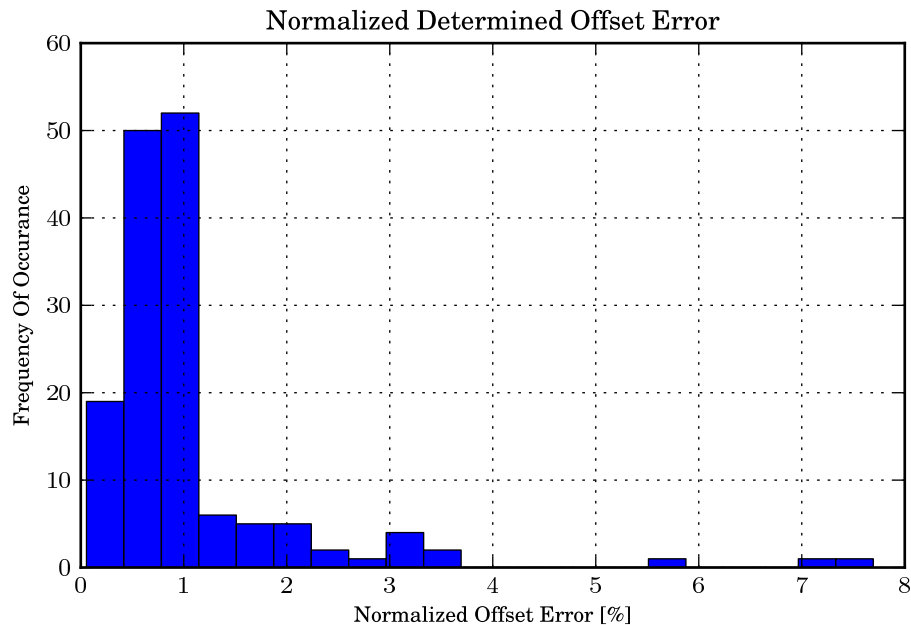
Figure E.44.:

To concisely represent the information displayed in figure Fig. E.40 for the numerous calibrations, a concise metric was made. This metric is long-winded but self-explanitory, its the mean of the magntiudes of the error in the embedding network's scatter parameters. A histogram of this metric for both the least squares calibration and the self-calibration are shown below. It shows that the error in self-calibration was smaller on average, by an order of magnitude.
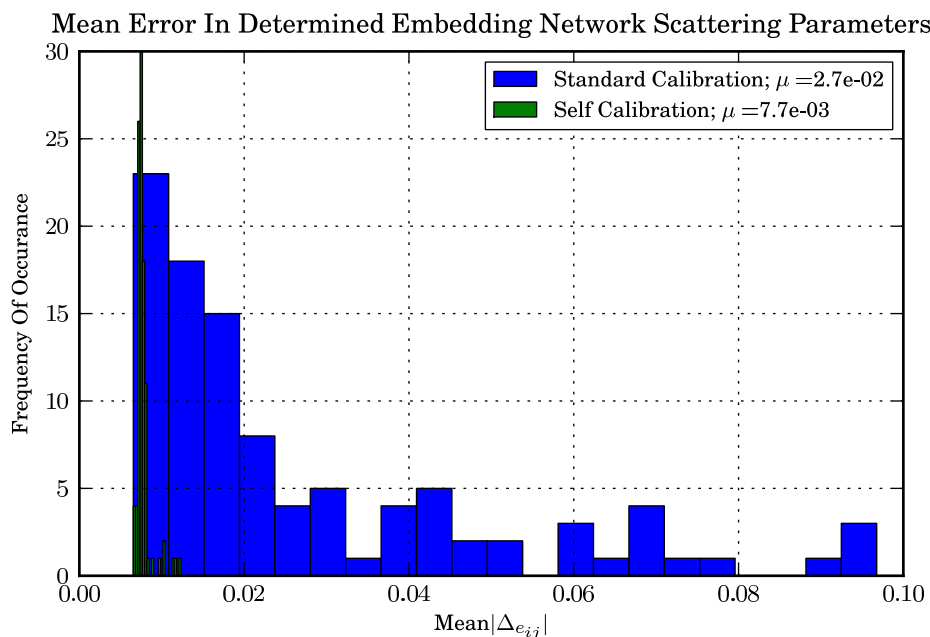
Figure E.45.:

E.3.5. Conclusion

Given the assumed models for systematic and random errors the self-calibration method can provide significant improvements in measurements, as well as provide a way to measure flange-aligment using standards which use the flanges themselves.

It is important to note that the simulation here assumes accurate knowledge of two other standards. In this case, those standards where a short and radiating open. This simulation also assumes there is no other source of systematic error.

E.3.6. Furture

Other calibration kits could be simulated such as, short, 2-delays, and a match. Also other sources of systematic errors could be introduced, such as waveguide dimensional uncertainty.

Bibliography

[1] Z. Popovic and E.N. Grossman. Thz metrology and instrumentation. *Terahertz Science and Technology, IEEE Transactions on*, 1(1):133 –144, sept. 2011.

[2] Huilin Li, A.R. Kerr, J.L. Hesler, Haiyong Xu, and R.M. Weikle. A ring-centered waveguide flange for millimeter- and submillimeter-wave applications. In *Microwave Symposium Digest (MTT), 2010 IEEE MTT-S International*, pages 604 –607, May 2010.

[3] Jeffrey L. Hesler Huilin Li, Anthony R. Kerr and Robert M. Weikle II. Repeatability and mismatch of waveguide flanges in the 500-750 GHz band. *Microwave Theory and Techniques, IEEE Transactions on*, to-be-published.

[4] T. J. Reck, L. Chen, C. Zhang, A. Arsenovic, C. Groppi, A. W. Lichtenberger, R. M. Weikle, and N. S. Barker. Micromachined probes for submillimeter-wave on-wafer measurements part i: Mechanical design and characterization. *Terahertz Science and Technology, IEEE Transactions on*, 1(2):349 –356, Nov. 2011.

[5] T. J. Reck, L. Chen, C. Zhang, A. Arsenovic, C. Groppi, A. Lichtenberger, R. M. Weikle, and N. S. Barker. Micromachined probes for submillimeter-

wave on-wafer measurements part ii: Rf design and characterization. *Terahertz Science and Technology, IEEE Transactions on*, 1(2):357 –363, Nov. 2011.

[6] D. F. Williams. 500 GHz - 750 GHz rectangular-waveguide vector-network-analyzer calibrations. *Terahertz Science and Technology, IEEE Transactions on*, 1(2):364 –377, Nov. 2011.

[7] Zhiyang Liu and R.M. Weikle. A reflectometer calibration method resistant to waveguide flange misalignment. *Microwave Theory and Techniques, IEEE Transactions on*, 54(6):2447 –2452, June 2006.

[8] Christian Perwass. *Geometric Algebra with Applications in Engineering.* Springer, softcover reprint of hardcover 1st ed. 2009 edition, November 2010.

[9] T.W. Crowe, B. Foley, S. Durant, Kai Hui, Y. Duan, and J.L. Hesler. Vna frequency extenders to 1.1 thz. In *Infrared, Millimeter and Terahertz Waves (IRMMW-THz), 2011 36th International Conference on*, page 1, oct. 2011.

[10] W.Grammer J.L.Hesler, A.R.Kerr and E.Wollack. In *Proc. 18'th International Symposium on Space Terahertz Technology*, pages 100 – 103, March 2007.

[11] Alexander Arsenovic and R.M. Weikle. Comparison of competing designs for delay-short calibration standards at wr-1.5. *International Conference on Infrared, Millimeter, and Terahertz Waves*, September 2012.

[12] N. Ridler. IEEE begins work on waveguide standards above 110 GHz [MTT-S society news]. *IEEE Microwave Magazine*, 10(1):122, February 2009.

[13] D. F. Williams, M. T. Ghasr, B. Alpert, Z. Shen, A. Arsenovic, R. M. Weikle, and R. Zoughi. Legendre fit to the reflection coefficient of a radiating rectangular waveguide aperture. *Antennas and Propagation, IEEE Transactions on*, 60(8):4009 –4014, aug. 2012.

[14] G.F. Engen. The six-port reflectometer: An alternative network analyzer. *IEEE*

Transactions on Microwave Theory and Techniques, 25(12):1075 – 1080, December 1977.

[15] Jr. Bauer, R.F. and P. Penfield. De-embedding and unterminating. *Microwave Theory and Techniques, IEEE Transactions on*, 22(3):282 – 288, mar. 1974.

[16] Robert E. Collin. *Foundations for Microwave Engineering - 2nd edition*. Wiley-IEEE Press, 2 edition, December 2000.

[17] S.J. Mason. Feedback theory-some properties of signal flow graphs. *Proceedings of the IRE*, 41(9):1144 –1156, September 1953.

[18] Moore-Penrose pseudoinverse, October 2012. Page Version ID: 517401172.

[19] D. Blackham. Application of weighted least squares to osl vector error correction. pages 11 – 21, jun. 2003.

[20] K. Wong. Uncertainty analysis of the weighted least squares vna calibration. In *ARFTG Microwave Measurements Conference, Fall 2004. 64th*, pages 23 – 31, 2-3 2004.

[21] Scattering parameters, October 2012. Page Version ID: 505163674.

[22] E. Bolinder. Clifford algebra: What is it? *Antennas and Propagation Society Newsletter, IEEE*, 29(4):18 – 23, August 1987.

[23] R.A. Speciale and N.R. Franzen. Super - tsd, a generalization of the tsd network analyzer calibration procedure, covering n - port measurements with leakage. In *Microwave Symposium Digest, 1977 IEEE MTT-S International*, pages 114 –117, 1977.

[24] H.-J. Eul and B. Schiek. A generalized theory and new calibration procedures for network analyzer self-calibration. *Microwave Theory and Techniques, IEEE Transactions on*, 39(4):724 –731, apr. 1991.

[25] Roger B. Marks. Formulations of the basic vector network analyzer error model including switch-terms. In *ARFTG Conference Digest-Fall, 50th*, volume 32, pages 115 –126, Dec. 1997.

[26] P.R. Young. Analysing connector repeatability of microwave vector measurements. In *Interconnections from DC to Microwaves (Ref. No. 1999/019), IEE Colloquium on*, pages 8/1 –8/5, 1999.

[27] A. Lewandowski and D. Williams. Stochastic modeling of coaxial-connector repeatability errors. In *Microwave Measurement Symposium, 2009 74th ARFTG*, pages 1 –4, 30 2009-Dec. 4 2009.

[28] M.J. Salter, N.M. Ridler, and P.M. Harris. Over-determined calibration schemes for rf network analysers employing generalised distance regression. In *ARFTG Microwave Measurements Conference, 2003. Fall 2003. 62nd*, pages 127 – 142, Dec. 2003.

[29] D. Williams. De-embedding and unterminating microwave fixtures with nonlinear least squares. *Microwave Theory and Techniques, IEEE Transactions on*, 38(6):787 –791, jun 1990.

[30] D.F. Williams, J.C.M. Wang, and U. Arz. An optimal vector-network-analyzer calibration algorithm. *Microwave Theory and Techniques, IEEE Transactions on*, 51(12):2391 – 2401, Dec. 2003.

[31] Linear least squares (mathematics), October 2012. Page Version ID: 516249612.

[32] D. Hestenes. *Space-Time Algebra*. Routledge, 1966.

[33] Tristan Needham. *Visual Complex Analysis*. Oxford University Press, USA, 1999.

[34] David Hestenes. The design of linear algebra and geometry. *Acta Applicandae Mathematicae*, 23:65–93, 1991.

[35] Chris Doran and Anthony Lasenby. *Geometric Algebra for Physicists*. Cambridge University Press, 2007.

[36] Homogeneous coordinates, October 2012. Page Version ID: 497760419.

[37] Arkadiusz Lewandowski, Wojciech Wiatr, and Janusz Dobrowolski. Multifrequency approach to the coaxial multiline through-reflect-line calibration. In *Microwave Radar and Wireless Communications (MIKON), 2010 18th International Conference on*, pages 1 –4, 2010.

[38] Alex Arsenovic et al. scikit-rf: Open-source microwave/rf tools for Python, 2009-present. http://www.scikit-rf.org.

[39] A. Rumiantsev and N. Ridler. Vna calibration. *Microwave Magazine, IEEE*, 9(3):86 –99, june 2008.

[40] G.F. Engen and C.A. Hoer. Thru-reflect-line: An improved technique for calibrating the dual six-port automatic network analyzer. *Microwave Theory and Techniques, IEEE Transactions on*, 27(12):987 – 993, Dec 1979.

[41] Norman R. Franzen and Ross A. Speciale. A new procedure for system calibration and error removal in automated s-parameter measurements. In *Microwave Conference, 1975. 5th European*, pages 69 –73, September 1975.

[42] R.B. Marks. A multiline method of network analyzer calibration. *Microwave Theory and Techniques, IEEE Transactions on*, 39(7):1205 – 1215, July 1991.

[43] W. Sigg and J. Simon. Reflectometer calibration using load, short and offset shorts with unknown phase. *Electronics Letters*, 27(18):1650 –1651, aug. 1991.

[44] G. Vandersteen, Y. Rolain, J. Schoukens, and A. Verschueren. An improved sliding-load calibration procedure using a semiparametric circle-fitting procedure. *Microwave Theory and Techniques, IEEE Transactions on*, 45(7):1027 –1033, jul. 1997.

[45] A. Lewandowski, W. Wiatr, and D. Williams. Multi-frequency approach to vector-network-analyzer scattering-parameter measurements. In *Microwave Conference (EuMC), 2010 European*, pages 260 –263, 2010.

[46] Bianca WIll and I. Rolfes. Multi-frequency calibration algorithms based on reduced self-calibration methods. *Microwave Theory and Techniques, IEEE Transactions on*, submitted for publication.

[47] Wikipedia. Non-linear least squares — wikipedia, the free encyclopedia, 2010.

[48] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[49] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[50] N.M. Ridler and M.J. Salter. A generalised approach to the propagation of uncertainty in complex s-parameter measurements. In *ARFTG Microwave Measurements Conference, Fall 2004. 64th*, pages 1 – 14, Dec. 2004.

[51] B.D. Hall. A computational technique for evaluating and propagating the uncertainty of complex-valued quantities. In *ARFTG Conference Digest, Fall 2002. 60th*, pages 19 – 28, Dec. 2002.

[52] Dylan Williams et al. Nist microwave uncertianty framework (beta version), 2011.

[53] Philip R. Bevington. *Data Reduction and Error Analysis for The Physical Sciences.* McGraw-Hill Companies, 1992.

[54] Barry N. Taylor and Chris E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of nist measurement results, technical report tn1297. *National Institute of Standards and Technology, Gaithersburg, MD.*, 2001.

[55] Wikipedia. Touchstone file — wikipedia, the free encyclopedia, 2011. [Online; accessed 5-November-2011].

[56] Alexander Arsenovic. Comparison of competing designs for delay-short calibration standards at WR-1.5. In *International Conference on Infrared, Millimeter, and Terahertz Waves*, September 2012.

[57] S. Maas. Circuit simulation in the dark ages. *IEEE Microwave Magazine*, 13(4):89 –99, June 2012.

[58] R.A. York, R.C. Compton, M. Kim, S. Wedge, and D.B. Rutledge. An alternative approach for designing microwave circuits using a personal computer. In *Antennas and Propagation Society International Symposium, 1988. AP-S. Digest*, pages 6 –9 vol.1, jun 1988.

[59] R.C. Compton. Perspectives in microwave circuit analysis. In *Circuits and Systems, 1989., Proceedings of the 32nd Midwest Symposium on*, pages 716 –718 vol.2, aug 1989.

[60] The Python Community. Python: Programming language, 2001–.

[61] Eric Jones, Travis Oliphant, Pearu Peterson, et al. Numpy: Numerical computing for python, 2001–.

[62] J.D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science Engineering*, 9(3):90 –95, may-June 2007.

[63] J.S. Seabold and J. Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, 2010.

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn:

Machine Learning in Python . *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[65] scikit-image team. scikit-image: Image processing in Python, 2009-present. http://www.scikit-image.org.

[66] SciKits - scikits. http://scikits.appspot.com/scikits.

[67] Google code. http://code.google.com/.

[68] GitHub. https://github.com/.

[69] Lihan Chen, A. Arsenovic, J.R. Stanec, T.J. Reck, A.W. Lichtenberger, R.M. Weikle, and N.S. Barker. A micromachined terahertz waveguide 90deg twist. *Microwave and Wireless Components Letters, IEEE*, 21(5):234 –236, may 2011.

[70] Alexander Arsenovic, L. Chen, M. Bauwens, H. Li, N. S. Barker, and R. M. Weikle. An experimentally-based technique for calibration uncertainty analysis. *Microwave Theory and Techniques, IEEE Transactions on*, submitted for publication, April 15, 2012.

[71] Alexander Arsenovic, L. Chen, T. J. Reck, N. S. Barker, A. Lichtenberger, and R. M. Wiekle. Multi-frequency one-port self-calibration using parametric standards. *Microwave Theory and Techniques, IEEE Transactions on*, to-be-published.

[72] David Hestenes. In: J.s.r. chisholm/a.k. commons (eds.), clifford algebras and their applications in mathematical physics. reidel, dordrecht/boston (1986), a unified language for mathematics and physics.

[73] H.A. Wheeler. Reflection charts relating to impedance matching. *Microwave Theory and Techniques, IEEE Transactions on*, 32(9):1008 – 1021, sep 1984.

[74] E.F. Bolinder. Geometric-analytic theory of transition in electrical engineering. *Proceedings of the IRE*, 47(6):1124 –1129, 1959.

[75] D. Hestenes. *New Foundations for Classical Mechanics (Fundamental Theories of Physics).* Springer, 1999.

[76] John A. Vince. *Geometric Algebra for Computer Graphics.* Springer, 2008.