Applications of Vector Fitting in the Solution of Electromagnetic Field Interactions

by

Evan Richards

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved December 2014 by the
Graduate Supervisory Committee:

Rodolfo Diaz, Chair
Konstantinos Tsakalis
Rodrigo Platte

ARIZONA STATE UNIVERSITY

May 2014

ABSTRACT

Vector Fitting (VF) is a recent macromodeling method that has been popularized by its use in many commercial software for extracting equivalent circuit's of simulated networks. Specifically for material measurement applications, VF is shown to estimate either the permittivity or permeability of a multi-Debye material accurately, even when measured in the presence of noise and interferences caused by test setup imperfections.

A brief history and survey of methods utilizing VF for material measurement will be introduced in this work. It is shown how VF is useful for macromodeling dielectric materials after being measured with standard transmission line and free-space methods.

The sources of error in both an admittance tunnel test device and stripline resonant cavity test device are identified and VF is employed for correcting these errors. Full-wave simulations are performed to model the test setup imperfections and the sources of interference they cause are further verified in actual hardware measurements. An accurate macromodel is attained as long as the signal-to-interference-ratio (SIR) in the measurement is sufficiently high such that the Debye relaxations are observable in the data.

Finally, VF is applied for macromodeling the time history of the total fields scattering from a perfectly conducting wedge. This effort is an initial test to see if a time domain theory of diffraction exists, and if the diffraction coefficients may be exactly modeled with VF. This section concludes how VF is not only useful for applications in material measurement, but for the solution of modeling fields and interactions in general.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

A common engineering task is deducing an equivalent circuit that has electrical characteristics that match those of a system being analyzed. In some cases, the system may only be known through a series of hardware tests or simulations, limiting knowledge to the discretely sampled data points that have been measured. In other scenarios, there may already exist a numerical recipe for predicting the behavior of the system under analysis, but the process would be much faster if there were a theoretical equation that could be evaluated directly. In these situations, the equivalent electrical circuit, and thus circuit theory provide an efficient means for numerically analyzing the system. The key objective is then to extract the circuit parameters that form the model of the system. Using a parallel plate capacitor permittivity measurement as an example, the typical issues encountered in attempting to meet this objective are now discussed.

1.1 Multi-Debye Parameter Extraction with Noise

A capacitive plate measurement under ideal circumstances confines the electric field to pass uniformly through the sample wedged in between the plates (electric fields which fringe around the plates are neglected) as shown in Figure 1.1.

The material changes the capacitance of the sample holder by its complex relative permittivity, ϵ_r, and as a result the sample holder and material are lumped into a single impedance element of complex capacitance, $C(\omega)$, as shown on the left in Figure 1.1. This frequency dependent complex capacitor models the material's mechanisms of energy storage and loss.

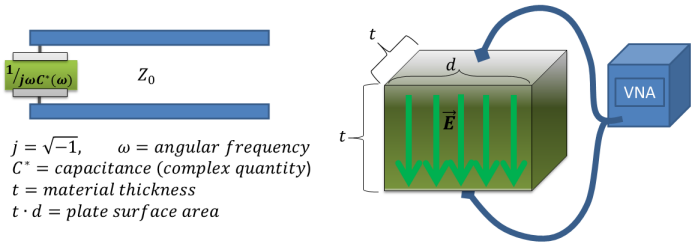

Figure 1.1: Equivalent transmission line of capacitive plate test device excited by a Vector Network Analyzer (VNA).

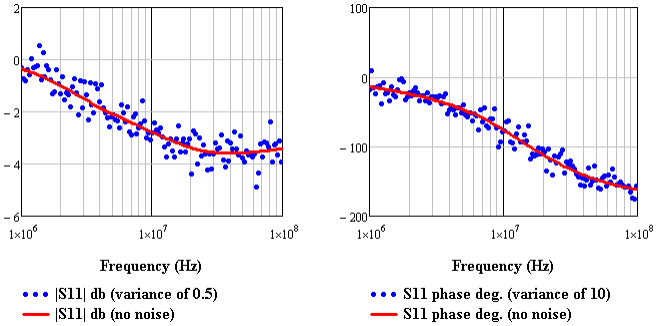

Figure 1.2: S_{11} measured from an assumed material impedance and with synthesized AWGN to simulate a real-life measurement scenario.

$$Z_{load} = \frac{1}{j\omega C} = \frac{1}{j\omega\epsilon_0\epsilon_r\frac{t\cdot d}{d}} \tag{1.1}$$

$$\epsilon_r = \frac{1}{Z_{load}j\omega\epsilon_0 t} \tag{1.2}$$

The measurement in Figure 1.2 is synthesized from Eq. (1.1) given an assumed permittivity consisting of a sum of Debye relaxation terms, as is typical for many RF materials. Additive White Gaussian Noise (AWGN) is then added to the reflection coefficient to simulate real-life thermal fluctuations and imperfections in the measured results.

Where the reflection coefficient, S_{11}, is related to the material impedance, Z_{load}, by:

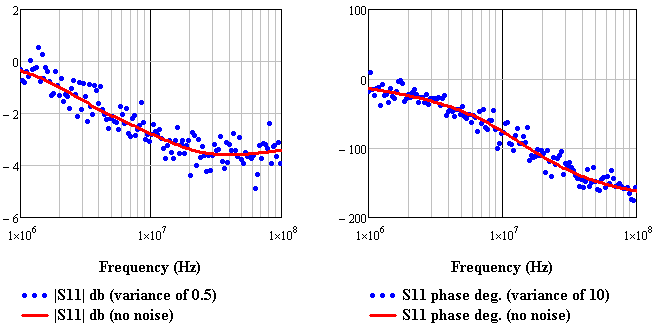

Figure 1.3: True and measured complex sample permittivity (with noise) inverted from Eq. (1.2).

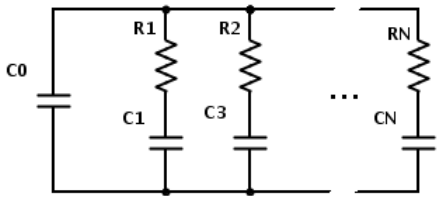

Figure 1.4: Equivalent circuit of measured multi-Debye material admittance.

$$Z_{load} = \frac{Z_0 \left(S_{11} + 1\right)}{1 - S_{11}} \tag{1.3}$$

where Z_0 is the characteristic impedance of the measurement cables and Vector Network Analyzer (VNA) that is the source of the test device. The measured impedance is then inverted to obtain the noisy measurement of the sample permittivity, $\epsilon_0 \epsilon_r^*$, using Eq. (1.2), where ϵ_0 is the free-space permittivity. The true (synthesized) and measured permittivity in the presence of noise are shown in Figure 1.3.

Under the multi-Debye assumption, the complex capacitance due only to the presence of the sample under test can be modeled by a ladder circuit, where each branch (as shown in Figure 1.4) dominates the material behavior over a finite frequency band about its relaxation frequency, $\frac{1}{RC}$. The C_0 term in Figure 1.4 models the ambient capacitance that is assumed constant over all frequencies (from the permittivity of the surrounding medium if not free space). The R_N term models the finite resistance from leakage current traversing microscopic conducting paths through the dielectric.

Modeling the measurement requires some approximation from at least the presence of noise corrupting the information in the measurement that is used for deriving a circuit such as in Figure 1.4. There is not a single circuit that will precisely characterize the measurement, but there is one that will with minimum error.

To find this circuit, the expression for the load impedance must be simplified so that the unknowns are linear quantities that may be solved for using one of the most elementary methods of approximating a function subject to noise: the method of lest squares. For a single resistor-capacitor branch (or RC branch), the load is rearranged as an admittance in Eq. (1.4).

$$Z_{load} = \frac{1}{j\omega C^*} = R + \frac{1}{j\omega C} = \frac{j\omega RC + 1}{j\omega C}$$

$$\frac{1}{Z_{load}} = Y_{load} = j\omega C^* = j\omega\epsilon_0\epsilon_r t = j\omega\left(\frac{C}{j\omega RC + 1}\right) \tag{1.4}$$

Normalizing to free-space and assuming the material thickness is one meter (so that t dissolves), the complex relative permittivity is given by Eq. (1.5).

$$\epsilon_r(s = j\omega) = \frac{\frac{C}{\epsilon_0}}{j\omega RC + 1} = \frac{\frac{1}{R\epsilon_0}}{s + \frac{1}{RC}} \tag{1.5}$$

Where $R_nC_n = \tau$, the relaxation time of a Debye function. From the viewpoint of admittance, a higher-order approximation of the measurement simply means the super-position of more RC branches expressed in the form of Eq. (1.5).

$$\epsilon_r = \frac{C_0}{\epsilon_0} + \sum_{n=1}^{N} \frac{\frac{1}{R_n\epsilon_0}}{s + \frac{1}{R_nC_n}} = d + \sum_{n=1}^{N} \frac{r_n}{s - p_n} \tag{1.6}$$

Given the permittivity measurement in Figure 1.3, the objective is to solve the unknowns, p_n, the pole singularities of the admittance, r_n, the residues of those singularities, and d, the constant term that represents the ambient or free-space asymptotic

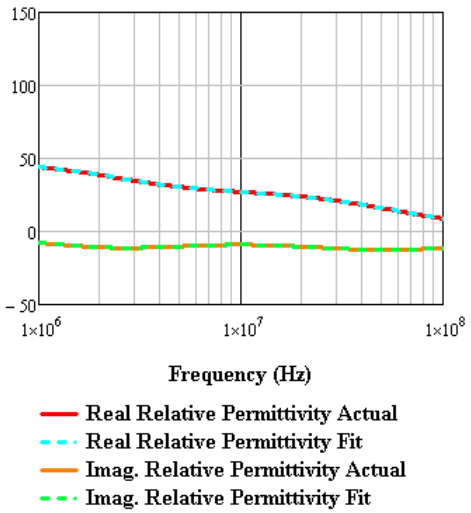

Figure 1.5: VF fit of complex permittivity without any noise present.

frequency value (e.g. the relative permittivity of free-space).

For the realistic case of a noisy measurement, the unknowns, r_n, p_n, d, can only be approximated. One of the recently popular methods for solving these coefficients is Vector Fitting (VF) [1] because it treats the non-linear problem as a two-stage linear problem. When noise is negligible, VF finds the exact poles and residues of the material to within machine precision (as shown in Figure 1.5) without any prior knowledge of the number or location of the poles or residues (i.e. the method can be implemented to adaptively refine the order of the approximation).

However, the solution may be biased by the noise when it is significant as seen in Figure 1.6.

Still, the generated fit for the example in Figure 1.6 is passive (has positive-valued residues) and has stable poles (negative-valued such that they exist on the left hand side of the complex frequency plane)[2], which means it provides an estimation that can be theoretically conceived from actual passive circuit components as in the schematic of Figure 1.4.

For the case of the measurement in Figure 1.6, the actual and fitted model are:

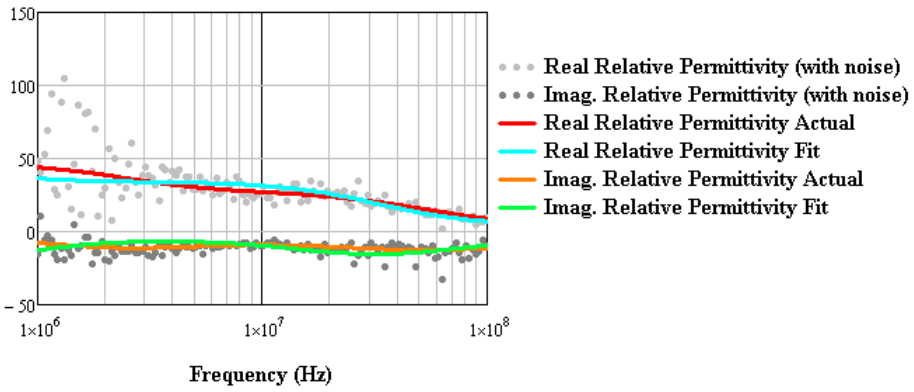

Figure 1.6: VF fit of complex permittivity with AWGN.

$$\epsilon_r^* actual - 1 = \frac{3E8}{s + 1.5E7} + \frac{1E9}{s + 2E8} + \frac{3E9}{s + 3E8} + \frac{2E9}{s + 5E8} + \frac{5E9}{s + 7E8}$$

$$\epsilon_r^* fit - 1 = \frac{11.8E9}{s + 3.8E9} + \frac{6.2E9}{s + 2.1E8} + \frac{1.4E6}{s + 7.9E7}$$

Where actual signifies the true, synthesized permittivity. The reduced order fit is a result of only the dominant poles in the topology of the permittivity being observable since the noise corrupts the more subtle fluctuations in the curve. Although the accuracy has diminished due to noise, this approximation provides a surrogate material which closely resembles the original material under test.

The first step towards manufacturing the material may be to form an equivalent circuit. This circuit could be built physically, or more commonly it would serve as a model for further simulations of the end-product design involving the material. The circuit is directly formed by the returned poles and residues from VF by converting them back to their circuit theory definition for the material's admittance:

$$R_n = \frac{1}{r_n \epsilon_0}, \qquad C_0 = d \cdot \epsilon_0 \qquad C_n = -\frac{1}{R_n p_n} \tag{1.7}$$

The circuit can then be incorporated into virtually any electromagnetic (EM)

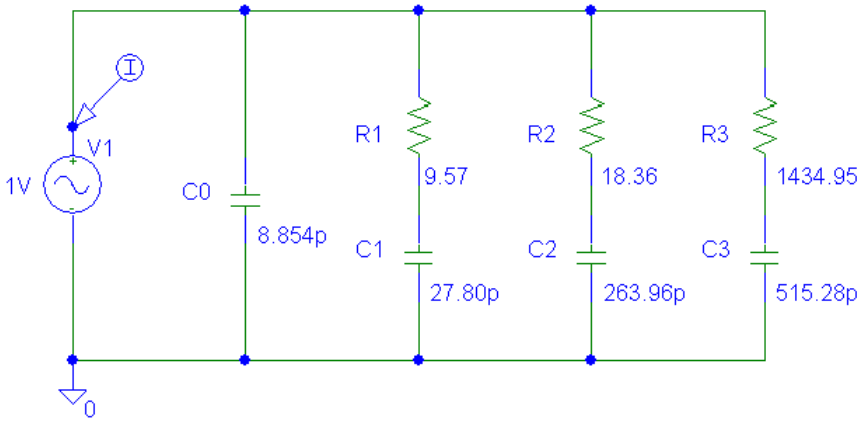

Figure 1.7: Generated circuit model derived from the VF result using Eq. (1.7).

simulation software, including the vast number of freely available circuit simulators such as PSpice.

For a unity frequency domain excitation voltage, the complex admittance of the material's equivalent circuit (shown in Figure 1.7) is simply $-I/V_1 = -I$ (the minus sign is an artifact of PSpice calculating current at a voltage source as flowing from the positive node to the negative node of the source so this needs to be reversed). This admittance agrees nicely with the admittance calculated directly from the estimated permittivity (the VF result) as shown in Figure 1.8.

The major advantage of the equivalent circuit form is that it may be more easily embedded in subsequent electromagnetic (EM) simulations (for instance geometrical surfaces may be given boundary conditions which obey circuit theory).

Furthermore, the equivalent circuit translates the frequency dependent behavior of the material into frequency independent circuit element values. Whereas frequency dependent material behavior must be deduced from observing a range of frequencies, frequency independent circuit elements may be compared directly. A circuit realization therefore also simplifies the design into a more concise description of the material.

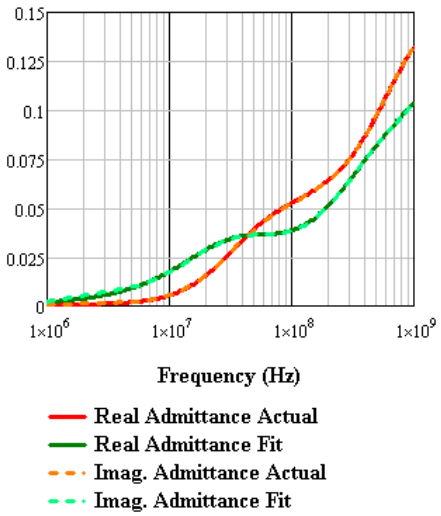

Figure 1.8: VF result of the admittance of a multi-Debye sample measured in the presence of noise. The VF result is computed using the analytical pole-residue model and using PSpice circuit simulation software.

1.2 Multi-Debye Parameter Extraction with Testing Imperfections

The capacitor measurement method can fail if it is extended into higher frequencies where its simplifying assumptions do not adequately account for all its "physics". For instance, when the excitation wavelength is twice the length of a structure, a wave may propagate along the test setup (such as from plate to plate or along the plates), creating standing waves that will distort the measurement. Without knowing the limitations of the test setup, there is no telling whether the apparent spurious results are setup related or material induced.

If these imperfections are known to occur at isolated frequencies (e.g. they are spurious resonances of the setup) they can be removed in most situations without any modifications to the test setup. The simplest approach may be to perform a measurement while excluding those frequencies and then use VF to form an analytic continuation between the measured bands (thus completing the entire spectrum of data).

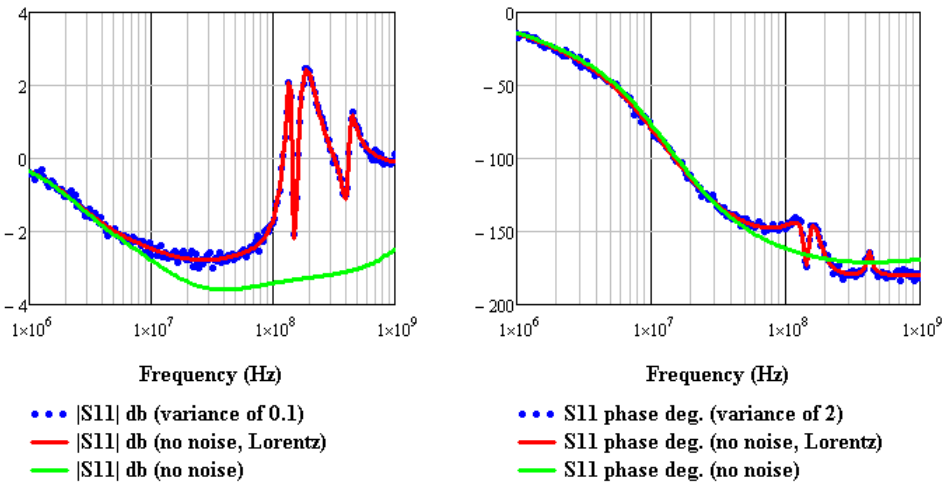

Figure 1.9: S11 (magnitude and phase) measurement of multi-Debye material without any interference (green), with spurious resonance interference (red), and with spurious resonances with noise (blue).

When invalid portions of the measurement cannot be easily circumvented, the next simplest approach is to penalize the faulty data points with a LS weight in the VF implementation. If that is not enough, a more rigorous method is to discriminate against certain complex pole locations of the vector fit result. For instance, if there is indeed a multi-Debye material under measurement, with no Lorentz-like behavior, any resonances observed must come from standing waves and so complex poles should be eliminated. This involves mathematically subtracting them from the model when found.

Consider the synthesized parallel plate measurement with broad frequency range in Figure 1.9. In this example, had the ideal circumstances assumed in the previous section still existed, the green line is what the measurement would truly have been. The red line is the true measurement when undesired standing waves form in the test setup (without any noise). The permittivity is then calculated as discussed in the previous section and is shown in Figure 1.10.

Constraining the pole locations in VF to be real-valued only (e.g. Debye terms)

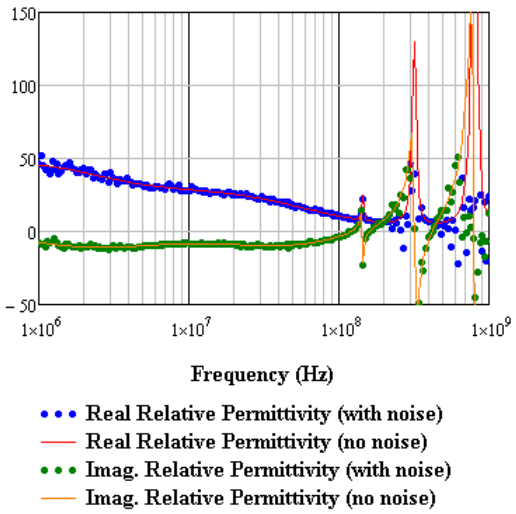

Figure 1.10: Real and imaginary relative components of multi-Debye permittivity in the presence of noise and interference (spurious resonances due to the test setup). The actual permittivity is indistinguishable above 100 MHz.

involves identifying all poles (real or complex) and then subtracting any complex poles out of the original measured data under analysis (so that they no longer bias the result).

When noise variance is negligible, the exact solution can be found just like the example given in the previous section. With significant noise variance, the actual complex poles are not precisely identified and this leads to the remaining Debye terms being inaccurately identified. This can be especially problematic when the complex poles have relatively large bandwidth such that the tail end of the curve on either side of the resonant frequency can be misinterpreted as Debye-like in shape.

When spurious poles have high enough quality factor (such that they occupy a narrow bandwidth and are hence isolated such that the multi-Debye poles are dominant) the complex poles are practically removed and a satisfactory fit may be generated as shown in Figure 1.11.

Notice that the complex pole which causes the lowest frequency resonance (near 100 MHz) has a wider bandwidth that may be interpreted as the sum of (valid) first

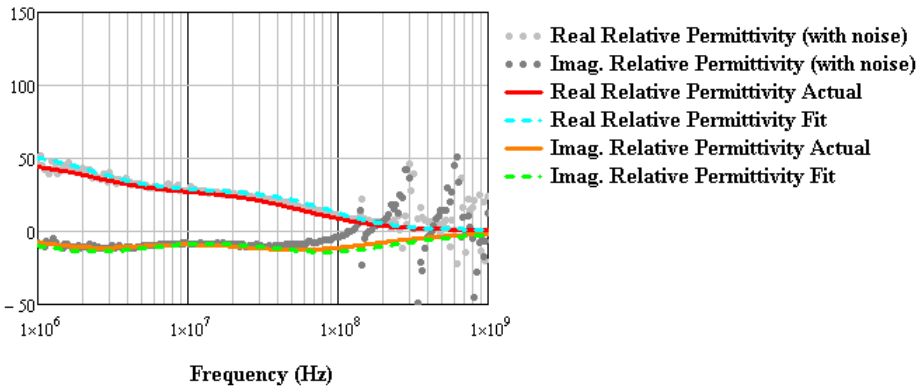

Figure 1.11: real and imaginary relative components of multi-Debye permittivity from Figure 1.10. A deterministic estimate is produced using VF.

order circuits. In some cases this can hinder the accuracy of the method when Lorentz terms are incorrectly approximated as Debye terms. This is especially true when the resonant frequencies of the complex poles are outside the measurement band, as they are no longer as observable. For instance, in this case, out of the 11 Debye and Lorentz terms, only 2 terms were necessary to accurately fit the data since most of the Debye relaxations were outside of the measurement band.

$$\epsilon_r^* fit - 1 = \frac{1.5E10}{s + 5.4E8} + \frac{3.8E8}{s + 1.4E7}$$

$$\begin{aligned}\epsilon_r^* actual - 1 = &\frac{3E8}{s + 1.5E7} + \frac{1E9}{s + 2E8} + \frac{3E9}{s + 3E8} + \frac{2E9}{s + 5E8} + \frac{5E9}{s + 7E8} \ldots \\ \ldots + &\frac{3E8 + 1jE7}{s + 7E6 + 9jE8} + \frac{3E8 - 1jE7}{s + 7E6 - 9jE8} + \frac{9E9 + 8jE7}{s + 7E7 + 2jE9} \ldots \\ \ldots + &\frac{9E9 - 8jE7}{s + 7E7 - 2jE9} + \frac{7E10 + 1jE8}{s + 2E8 + 5jE9} + \frac{7E10 - 1jE8}{s + 2E8 - 5jE9}\end{aligned}$$

1.3 Analytic Continuation

To identify a material's properties over a wide range of frequencies, often several individual experiments are performed, each involving a separate test device that may only operate over a finite band of frequencies. Due to device imperfections, these

separate experiments don't always give data that agree in the overlap region of their frequency ranges. Furthermore, there may be frequency bands inaccessible to convenient test methods. As such, there are gaps of missing frequency bands in the sum total measurement from all devices.

Rather than attempting to measure these missing regions by constructing new devices (if possible), it is much easier to interpolate the data between the sampled frequency bands. Since VF can provide a model for the data in the form of an analytic function, extending the observed frequency domain to fill these gaps can be done by analytic continuation.

The pole and residue models derived for the material permittivity in the previous sections are said to be analytic because their derivatives can be taken along any direction in the complex 2D plane and are defined at all points except at the location of the pole singularities (where in this region they are meromorphic functions). For instance, the derivative of Eq. (1.5) (where for simplicity it has been normalized by ϵ_0 in Eq. (1.8)) is defined everywhere in the complex plane except at $s_0 = \frac{1}{R_n C_n}$, which is evident in Eq. (1.8).

$$\frac{d}{ds}\left(\frac{\frac{1}{R_n}}{s+\frac{1}{R_nC_n}}\right) \quad \underset{\text{Reciprocal Law}}{\Rightarrow} \quad \frac{1}{R_n}\left(\frac{-\frac{d}{ds}\left(s+\frac{1}{R_nC_n}\right)}{\left(s+\frac{1}{R_nC_n}\right)^2}\right) = \frac{-\frac{1}{R_n\epsilon_0}}{\left(s+\frac{1}{R_nC_n}\right)^2} \tag{1.8}$$

However Eq. (1.8) does not check that the derivative is path independent. A more rigorous test that does is by applying the Cauchy-Riemann conditions,

$$f(s) = u(s) + jv(s) \quad \text{is analytic if} \quad \frac{\partial u}{\partial \sigma} = \frac{\partial v}{\partial \omega} \text{ and } \frac{\partial v}{\partial \sigma} = -\frac{\partial u}{\partial \omega}$$

For $s = \sigma + j\omega$, the real and imaginary components of Eq. (1.5), u, v, are

$$\frac{\frac{1}{R_n}}{(\sigma+j\omega)+\frac{1}{R_nC_n}} = \frac{\frac{1}{R_n}}{\frac{\sigma\cdot R_nC_n+1}{R_nC_n}+j\omega}$$

$$\begin{aligned} &= \frac{\frac{1}{R_n}}{a+jb} = \frac{\frac{a}{R_n}}{(a^2+b^2)} - j\frac{\frac{b}{R_n}}{(a^2+b^2)} = u(s)+jv(s) \\ &= \frac{C_n\left(\sigma R_nC_n+1\right)}{\sigma^2R_n^2C_n^2+2\sigma R_nC_n+\omega^2R_n^2C_n^2+1} - j\frac{\omega R_nC_n^2}{\sigma^2R_n^2C_n^2+2\sigma R_nC_n+\omega^2R_n^2C_n^2+1} \end{aligned} \tag{1.9}$$

Where the denominator terms in Eq. (1.9) are the same and for simplicity are renamed,

$$u(s)+jv(s) = \frac{C_n\left(\sigma R_nC_n^2+C_n\right)}{a+b+1} - j\frac{\omega\, R_nC_n^2}{a+b+1}$$

Where $a = \sigma^2R_n^2C_n^2+2\sigma R_nC_n$ and $b = \omega^2R_n^2C_n^2$. Using the chain rule and reciprocal law, the following partial derivatives are shown to be equal as expected for them to be analytic.

$$\begin{aligned} \frac{\partial u}{\partial\sigma} &= \frac{R_nC_n^2}{a+b+1} - \left(\sigma R_nC_n^2+C_n\right)\frac{\frac{\partial}{\partial\sigma}(a+b+1)}{(a+b+1)^2} \\ &= \frac{R_nC_n^2}{a+b+1} - \frac{2R_nC_n^2(a+1)}{(a+b+1)^2} = \frac{R_nC_n^2(a+b+1)}{(a+b+1)^2} - \frac{2R_nC_n^2(a+1)}{(a+b+1)^2} \\ &= \frac{R_nC_n^2(b-a-1)}{(a+b+1)^2} \\ \frac{\partial v}{\partial\omega} &= -\frac{R_nC_n^2}{a+b+1} + \left(\omega R_nC_n^2\right)\frac{\frac{\partial}{\partial\omega}(a+b+1)}{(a+b+1)^2} \\ &= -\frac{R_nC_n^2(a+b+1)}{(a+b+1)^2} + \frac{2R_nC_n^2b}{(a+b+1)^2} \\ &= \frac{R_nC_n^2(b-a-1)}{(a+b+1)^2} \end{aligned}$$

The cross terms are also equal in magnitude but vary by their sign as expected for them to be analytic.

$$\frac{\partial u}{\partial \omega} = -\frac{\left(\sigma R_n C_n^2 + C_n\right) 2\omega R_n^2 C_n^2}{\sigma^2 R_n^2 C_n^2 + 2\sigma R_n C_n + \omega^2 R_n^2 C_n^2 + 1}$$
$$= -\frac{2\sigma\omega R_n^3 C_n^4 + 2\omega R_n^2 C_n^3}{\sigma^2 R_n^2 C_n^2 + 2\sigma R_n C_n + \omega^2 R_n^2 C_n^2 + 1}$$

$$\frac{\partial v}{\partial \sigma} = \frac{\omega R_n C_n^2 \left(2\sigma R_n^2 C_n^2 + 2R_n C_n\right)}{\sigma^2 R_n^2 C_n^2 + 2\sigma R_n C_n + \omega^2 R_n^2 C_n^2 + 1}$$
$$= \frac{2\sigma\omega R_n^3 C_n^4 + 2\omega R_n^2 C_n^3}{\sigma^2 R_n^2 C_n^2 + 2\sigma R_n C_n + \omega^2 R_n^2 C_n^2 + 1}$$

Since the pole-residue basis that VF uses is analytic, interpolating its result to frequency points that are not known by measurement is still a valid estimate. Performing an analytic continuation on this data can be done without any modification to the VF algorithm.

Much like a single frequency band measurement is identified by VF, separate bands concatenated together may be fit (with a discontinuity in the sampling frequency where they are concatenated). For example, consider Figure 1.12, where a synthesized permittivity has been plotted with red and blue regions indicating where an actual measurement has been taken (due to the hypothetically finite operating frequency range of the available test methods).

In this example, only frequencies below 10 MHz and above 1GHz were able to be measured from two separate synthesized experiments. Below 10 MHz, the measurement would likely be made with a transmission line method (such as the parallel plate capacitor method discussed previously), but for the measurement above 1 GHz, a free-space method should be used to avoid spurious resonances and onset of higher order modes in the test setup.

After measuring both reflection and transmission coefficients in an ideal free-space measurement, the permittivity may be directly inverted from a Nicholson-Ross method[3, 4]. To simulate these two measurements, AWGN is added to the magnitude

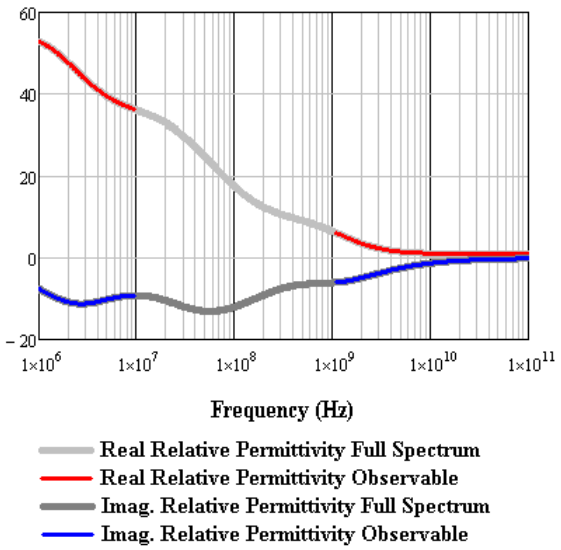

Figure 1.12: The true real and imaginary permittivity (shown in gray scale) and true real and imaginary permittivity that are being measured by the finite bandwidths of the test setups (blue and red regions of the curves).

and phase with variance 0.5 and 10 respectively for the lower frequency measurement and similarly with variance 2 and 1 for the higher frequency measurement. The two measured bands are plotted against the true permittivity in Figure 1.13.

In this case, the lower frequency spectrum of the measurement does not offer nearly as much information about the material as does the higher spectrum, and because the majority of the poles for the actual permittivity reside in the gap and lower spectrum, this is a particularly challenging curve fitting problem.

This challenge can be met, first, by assuming the permittivity should approach free-space at infinite frequency, for which VF can be constrained to force the constant term to equal $d = 1$. Second, the error tolerance may be relaxed since clearly the variance is far too high in the lower spectrum to achieve a precise fit. With these constraints and by constraining the residues to being passive[5, 6], a suitable, reduced-order fit can be found in an entirely automated process, with final results shown in Figure 1.14.

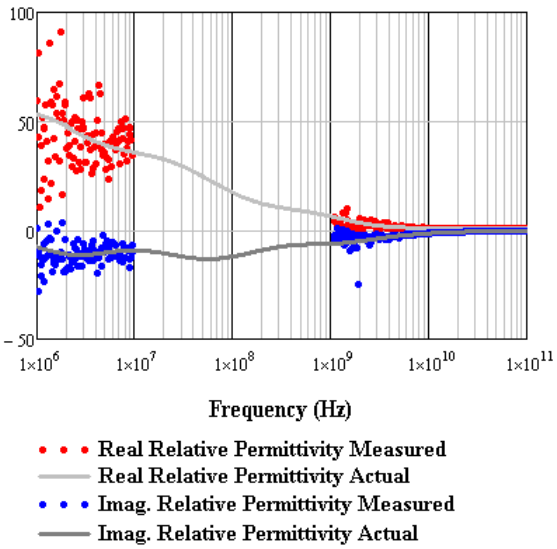

Figure 1.13: Measured bands of the real and imaginary components of the permittivity (red and blue). Gray scale signifies regions of the true permittivity that were not measured.

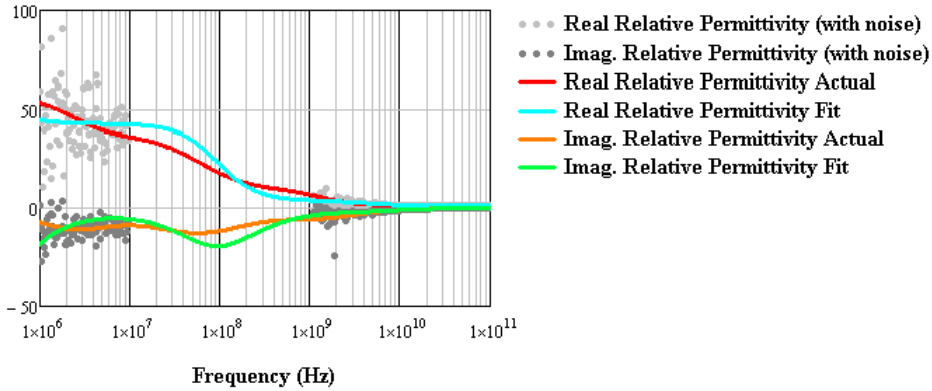

Figure 1.14: VF result of data measured in separate bands as indicated in 1.13.

This is perhaps a worst-case scenario for analytic continuation with VF since the system's true poles are located below the high frequency measurement where information is missing or highly corrupted by noise. The following reduced-order fit is

$$\epsilon_r^* actual - 1 = \frac{3E8}{s+1.5E7} + \frac{1E9}{s+2E8} + \frac{3E9}{s+3E8} + \frac{2E9}{s+5E8} + \frac{5E9}{s+7E8} + \frac{7E10}{s+8E9}$$

$$\epsilon_r^* fit - 1 = \frac{6.7E10}{s+2.7E10} + \frac{2.3E10}{s+6.0E8} + \frac{1.2E8}{s+6.6E5}$$

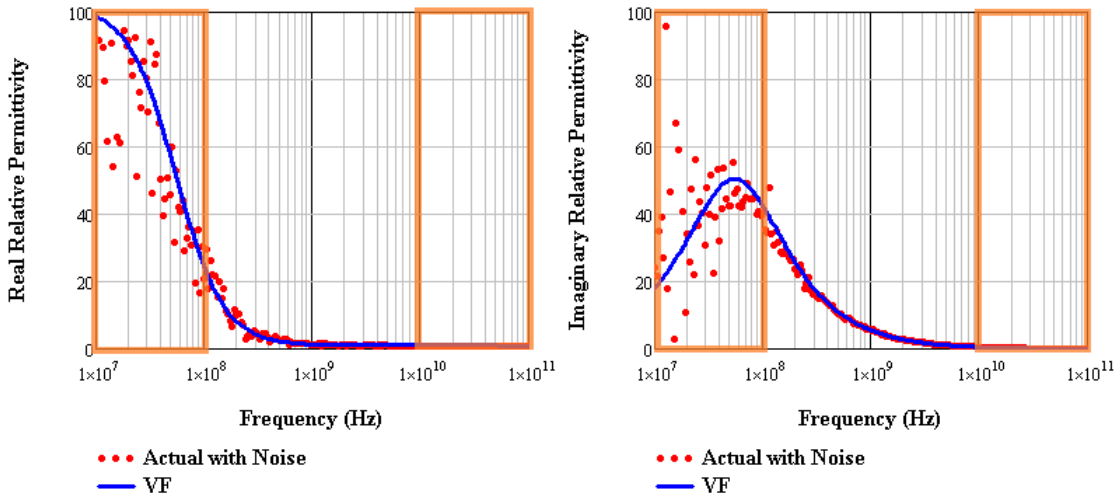

Figure 1.15: VF result of data measured in separate bands that are indicated by regions highlighted with orange bars.

Assuming a less noise-corrupted measurement can be attained, the results can be much more accurate. In Figure 1.15, the measured frequency bands are signified by the orange rectangular regions. Although the measured frequency bands are still separated by a large missing band, the noise power is reasonable and so VF finds an appropriate fit.

For this particular case, the Debye relaxations are roughly within 10 MHz to 100 MHz measured band and so they are observable and therefore are curve fit accurately as shown in Figure 1.16. The relaxation frequencies are indicated in Figure 1.16 by green X marks and the located relaxation frequencies (by VF) are shown by violet plus signs.

However, if the poles are located outside the measurement band, the fit can be more challenging for VF. This is shown in Figure 1.17 and Figure 1.18 where, given the same amount of noise, the fit is much less accurate since the Debye relaxations were not nearly as observable as in the previous case.

Therefore, for accurate analytic continuation results, several experiments should be conducted such that most of the Debye relaxations are measured. Then, these mea-

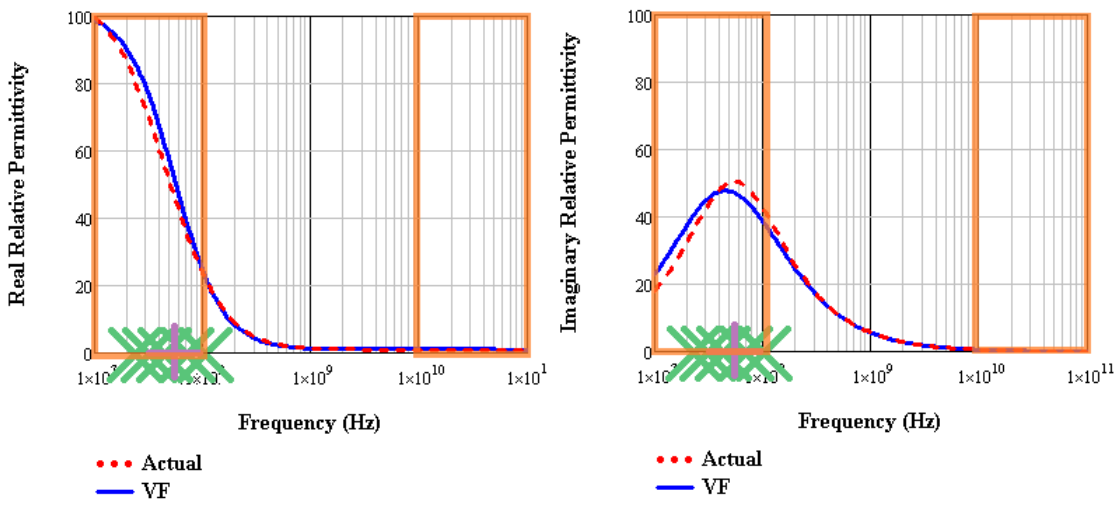

Figure 1.16: VF result of data measured in separate bands that are indicated by regions highlighted with orange bars. The true signal is shown along with the actual pole locations (green X marks). The located poles are shown as blue plus signs).

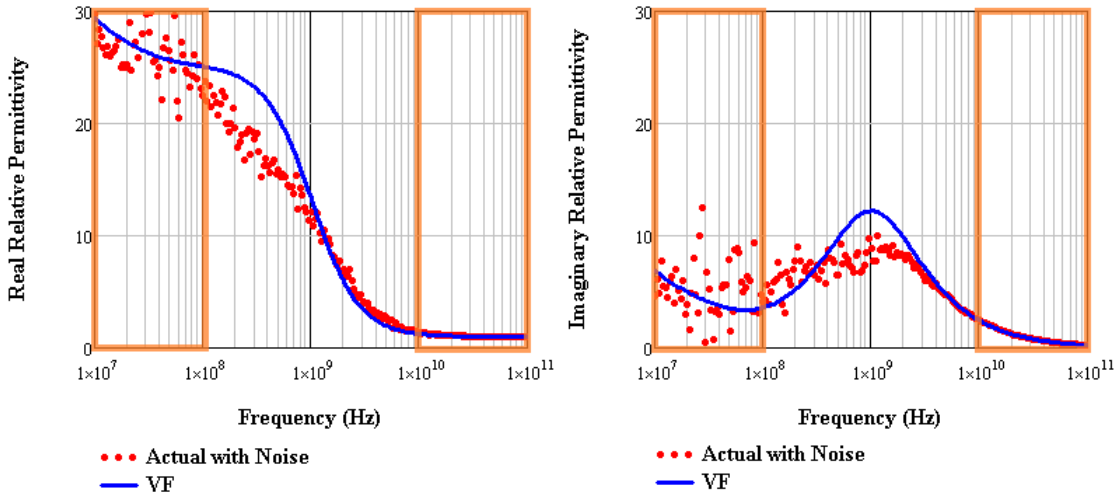

Figure 1.17: VF result of data measured in separate bands that are indicated by regions highlighted with orange bars.

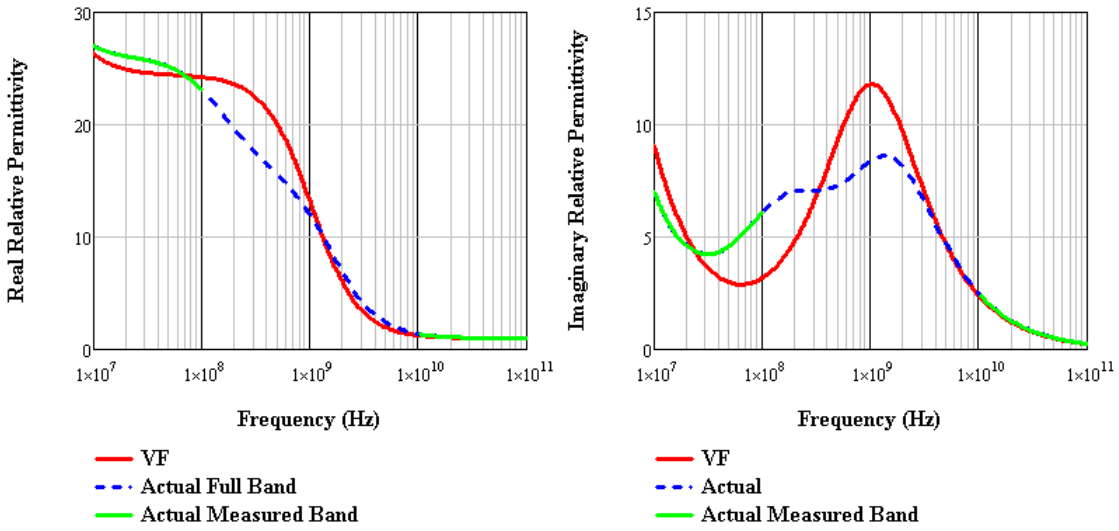

Figure 1.18: VF result of data measured in separate bands that are indicated by regions highlighted with orange bars. The true signal is shown and all pole locations are outside the measured band.

surements may be *stitched* together by literally vertically concatenating the measured data and performing the regular VF procedure.

Chapter 2

VECTOR FITTING ALGORITHM

2.1 State Space Formulation of the Problem

Vector Fitting (VF) is a method useful for approximating the state equations of a system by accurately locating its dominant poles and residues over an observed domain. The name "Vector Fitting" stems from how the method may be easily generalized to simultaneously fit a vector of system transfer functions (vector transfer function)[1]. The method is commonly applied to the frequency domain where a scalar or vector transfer function is the assumed observation to be fitted to (or an impulse response in the time domain and similarly for other such analogous domains).

VF is a State Equation Approximation (SEA) and therefore time and frequency domain approximations are parameterized by the exact same solution of state matrices. The only difference computationally is in the basis used during the least squares stages of the algorithm. This introduction focuses on defining the problem's state space realization (SSR) since individual time and frequency (or other) domain implementations are easily understood as extensions of this problem.

VF approximates a linear time invariant (LTI) system's poles and residues (linear and time invariant such that the SSR is a superposition of scalar state matrices – no time dependency). In the frequency domain, the system's transfer function, $H(s)$, can then be modeled as a sum of the residues, r_n, over first-order poles, p_n:

$$H(s) = \sum_{n=0}^{N} \frac{r_n}{s - p_n} \tag{2.1}$$

Where $s = \sigma + j\omega$, is the complex frequency. In the time domain this corresponds

to an impulse response that is a superposition of damped sinusoids:

$$h(t) = \sum_{n=0}^{N} r_n exp(p_n t) \tag{2.2}$$

The information in both Eq. (2.1) and Eq. (2.2) can be compacted into their SSR. In the time domain this is:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{2.3}$$

In the frequency domain the state matrices: A, B, C, D, still store all the vital information about the system, which is found by taking the Laplace transform of Eq. (2.3),

$$\begin{aligned} L\{\dot{x}(t)\} &= sX(s) = AX(s) + BU(s) \\ L\{y(t)\} &= Y(s) = CX(s) + DU(s) \end{aligned} \tag{2.4}$$

and then factoring the state vector, $X(s)$,

$$\begin{aligned} sX(s) &= AX(s) + BU(s) \\ sX(s) - AX(s) &= BU(s) \\ (sI - A)X(s) &= BU(s) \\ X(s) &= (sI - A)^{-1}BU(s) \end{aligned} \tag{2.5}$$

$H(s)$ is found by substituting Eq. (2.5) into the output, $Y(s)$ of Eq. (2.4), and then dividing by the input, $U(s)$,

$$\begin{aligned} Y(s) &= CX(s) + DU(s) \\ Y(s) &= C(sI - A)^{-1}BU(s) + DU(s) \\ H(s) = \frac{Y(s)}{U(s)} &= C(sI - A)^{-1}B + D \end{aligned} \tag{2.6}$$

In Eq. (2.6) the r_n and p_n terms have been collected into the C and A matrices where C is a column of the N residues, I is an $N \times N$ identity matrix, A is an $N \times N$ diagonal state matrix containing the poles, B is the $N \times 1$ frequency domain source spectrum (for simplicity assume an impulse source such that it is unity for all partial fraction terms at all frequencies), and D is a 1×1 (or constant) direct transmission element that may also be thought of as the asymptotic value of $H(s)$ at infinite frequency when all other terms approach zero. Then, Eq. (2.6) is equivalently written in its explicit matrix form, where d is the asymptotic constant term (and is zero for this case).

$$H(s) = \sum_{n=0}^{N} \frac{r_n}{s - p_n} = \begin{bmatrix} r_0 \\ \vdots \\ r_N \end{bmatrix}^T (sI - A)^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{N \times 1} + [d]$$

where,

$$sI - A = s \begin{bmatrix} 1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1 \end{bmatrix}_{N \times N} - \begin{bmatrix} p_0 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & p_N \end{bmatrix}$$

Whereas Eq. (2.1) and Eq. (2.2) express the poles of the system as non-linear (denominator) quantities, the system's SSR represents all pertinent quantities in the two linear equations of Eq. (2.3). As will be shown, this permits a linear solution of the system's poles. The SSR also generalizes the solution for both time and frequency (or conceivably other) domains. That is, the solution of the system's poles is an identical process in either domain. The residue terms are linear in Eq. (2.1) and Eq. (2.2) and can be solved as a simple least squares solution after the poles are identified. This least squares process will be discussed in the following sections.

2.2 Pole Relocation

The first step in the VF algorithm is to refine the best guess of the location of the system's poles. This implies that at start, the user initially guesses the quantity and location of poles, p_n, which are usually assumed linearly spaced or logarithmically spaced throughout the frequency band of the observation (complex starting poles may also be assumed as spaced along the imaginary axis with small real components). VF then iteratively refines this initial guess or *relocates* the poles until the modeled transfer function approximates the observed system to within an acceptable error tolerance.

Generally for other methods, this problem is stated as non-linear with unknown coefficients for the numerator, N, and denominator, D, polynomials of H. With these polynomials represented in factored form (such that H has zeros in the numerator and poles in the denominator):

$$H(s) = \frac{N(s)}{D(s)} = \frac{n_p s^p + \ldots + n_1 s + n_0}{d_q s^q + \ldots + d_1 s + d_0} \tag{2.7}$$

The solution for n_p cannot be found without prior knowledge of d_q. However, a simplification to this problem exists by considering the concept of zeros and poles as fundamentally the same; they are both roots of their respective polynomials in Eq. (2.7). Exploiting this similarity, the denominator poles can be viewed as zeros, z'_n, after dividing both the numerator and denominator by the set of initially guessed poles (such that they cancel by division and do not change the problem).

$$H(s) = \frac{N(s)}{D(s)} = \frac{\prod_n (s - z_n)}{\prod_n (s - p_n)} = \frac{\frac{\prod_n (s-z_n)}{\prod_n (s-p_n)}}{\frac{\prod_n (s-z'_n)}{\prod_n (s-p_n)}} = \frac{d + \sum_n \frac{r_n}{s+p_n}}{(1+d') + \sum_n \frac{r'_n}{s+p_n}} \tag{2.8}$$

Finally, the numerator and denominator in Eq. (2.7) are expanded into sums of

residues over first-order poles such that all unknowns now appear as numerator terms. The relation between the true transfer function, H, and the fitted residues, $[r_n, r'_n]$, poles, $[p_n, p'_n]$, and constant terms, $[d, d']$, is then linearized to:

$$d + \sum_n \frac{r_n}{s + p_n} = H(s) \left((1 + d') + \sum_n \frac{r'_n}{s + p_n} \right) \tag{2.9}$$

As the solution converges, the right hand side of Eq. (2.9) approaches H. From this perspective, the primed variables may be conceptualized as error terms that approach zero upon convergence. Another perspective, from the original authors of the method, is that H is scaled by a "smoothing" or "molding" function, σ[7]. This is why VF was originally titled "Vector Fitting with Optimal Scaling".

$$\sigma(s) = \left((1 + d') + \sum_n \frac{r'_n}{s + p_n} \right) \tag{2.10}$$

σ then molds (as in physically modifies the shape of the curve) in the sense that it weights the true system by the assumption made about the currently estimated poles. Without σ, the least squares solution for the residues would be ruinously biased by the trivial assumption of initially guessed poles (unless it was a very well-informed initial guess).

If σ were converted back to pole-zero form, according to Eq. (2.8) the relocated poles are then the roots of the denominator polynomial, D. Rather than implement a numeric root solver (such as Newton's method, bi-section, etc.) to find the roots of D, the SSR allows the new poles to be realized directly from the solved residues, r'_n, and current pole estimates, p'_n, by stating it as an eigenvalue problem.

Updating the current estimate of p_n with the SSR approach can be conceptually thought of as inverting σ such that z'_n now appear in the denominator (and are then effectively poles). This is equivalent to interchanging the input, Y, with the output,

U, such that $\sigma = Y/U$ becomes $1/\sigma = U/Y$. Therefore, using the SSR this can be achieved directly without explicitly realizing the pole-zero form of Eq. (2.8). Consider the state equations for (in arbitrarily the time or frequency domain) after interchanging input for output (Eq. (2.3) which is now shown here again for convenience).

$$\dot{x}(t) = Ax(t) + By(t)$$

$$y(t) = Cx(t) + Dy(t)$$

the second expression in Eq. (2.3) is then solved for in terms of y,

$$\begin{aligned} u(t) &= C \cdot x(t) + D \cdot y(t) \\ u(t) - C \cdot x(t) &= D \cdot y(t) \\ D^{-1}\left(u(t) - C \cdot x(t)\right) &= y(t) \end{aligned} \tag{2.11}$$

This is the expression for the output of the inverted transfer function. Plugging this new expression for y into the differential equation of Eq. (2.3) and then collecting the C matrix (containing r'_n) with the A matrix (containing p_n),

$$\begin{aligned} \dot{x}(t) &= A \cdot x(t) + B \cdot D^{-1}\left[u(t) - C \cdot x(t)\right] \\ \dot{x}(t) &= A \cdot x(t) + B \cdot D^{-1}\left[u(t) - C \cdot x(t)\right] \\ \dot{x}(t) &= \left(A - B \cdot D^{-1}C\right) \cdot x(t) + B \cdot D^{-1} \cdot u(t) \end{aligned} \tag{2.12}$$

Although $A = \mathrm{diag}(p_n)$, the $N \times N$ matrix $B \cdot D^{-1}C$ is not diagonal and therefore the new poles (or eigenvalues) must be solved as the roots of $\det[sI - (A - BD^{-1}C)]$ from the highlighted expression in Eq. (2.12). As discussed earlier, D contains only a constant term, $[d]$, C is a column vector of residues, B is a column vector of ones, and in addition if the eigenvalue solution process is lumped into a single function, eig, we get that the relocated poles are

$$p_n = eig[A - BC^T/d]$$

.

Solving the roots of a polynomial as an eigenvalue problem is a robust approach and is not an uncommon practice (e.g. Matlab's *roots* function solves the roots of polynomials using an eigenvalue method). After a single iteration of relocating has been completed, the current estimate of the poles is updated with these new locations. It is then on to the second stage of the fitting process where the residues are identified. This concludes the pole relocation process. In summary, the pole-relocation process is a linear solution that is made possible by assuming a basis consisting of partial fractions. This is ensued by dividing the numerator and denominator of the rational polynomial approximation by a common set of poles that does not change the problem but allows it to be reformulated such that all unknowns are linear. The SSR can then be manipulated to obtain a linear expression for the relocated poles.

2.3 Least Squares Problem

When identifying an LTI system, VF may be correctly applied when the linear residue terms in Eq. (2.1) and Eq. (2.2) can be uniquely identified as coefficients in a traditional least squares problem: $Ax = b$. Here, the A matrix has columns comprised of the basis functions of the observation (e.g. transfer function or impulse response) in its respective domain, with residue terms stored separately in the x vector (as coefficients for each column of A). The fit then minimizes the L2-norm squared error, $|Ax - b|_2^2$ where b is the observation that is meant to be identified analytically.

As long as y may be uniquely determined in the least squares problem, the VF solution may apply. For instance, VF may not yield a sufficiently accurate fit when the observation contains unknowns in addition to the modeled system's poles and residues. A common example of this would be when the observation contains an unknown constant delay or frequency dependent time delay. Before VF may apply,

these additional effects must be modeled and divided or subtracted out from the observation such that the only unknowns are the modeled system's poles and residues.

Assuming the LTI system can be modeled similarly as Eq. (2.1) and Eq. (2.2), with the only unknowns being the residues and poles, then the least squares problem may be defined as follows. Assume the user has guessed an initial set of poles as in Section 3.2. The basis of the normal equations must then be derived. In the frequency domain, Eq. (2.9) can be formulated into a system matrix of the least squares normal equations by rearranging the observation, H, to one side.

$$\begin{aligned} d + \sum_n \frac{r_n}{s+p_n} &= H(s)\left((1+d') + \sum_n \frac{r'_n}{s+p_n}\right) \\ d + \sum_n \frac{r_n}{s+p_n} - H(s)\cdot\sum_n \frac{r'_n}{s+p_n} - H(s)\cdot d' &= H(s) \end{aligned} \tag{2.13}$$

Because the rows of the system matrix, A, vary only as a function of the complex frequency, s, only a single row is shown for succinctness. Then, least squares problem is stated in Eq. (2.14) assuming $d' = 0$ (for the original VF method).

$$\underbrace{\left[\begin{array}{cccccc} \frac{1}{s-p_0} & \dots \frac{1}{s-p_n} & 1 & -\frac{H(s)}{s-p_0} & \dots -\frac{H(s)}{s-p_N} & -H(s) \end{array}\right]}_{A_{0,:}} \cdot \underbrace{\left[\begin{array}{c} r_0 \\ \dots r_N \\ d \\ r'_0 \\ \dots r'_N \\ d'=0 \end{array}\right]}_{x} = \underbrace{H(s)}_{b} \tag{2.14}$$

The time domain implementation of VF is found from the inverse Laplace transform of the frequency domain basis. First, H is separated into input and output

components so as to not restrict the time domain basis to only impulse response characterization (i.e. so that common sources like unit step or impulse inputs may be applied).

$$\begin{aligned} Y(s)\left((1+d')+\sum_n \frac{r'_n}{s+p_n}\right) &= U(s)\left(d+\sum_n \frac{r_n}{s+p_n}\right) \\ Y(s) = U(s)\sum_n \frac{r_n}{s+p_n} + U(s)\cdot d - Y(s)\cdot\sum_n \frac{r'_n}{s+p_n} &- Y(s)\cdot d' \end{aligned} \tag{2.15}$$

Then the inverse Laplace transform of Eq. (2.15) is,

$$\mathcal{L}\{Y\}^{-1} = y(t) = \sum_n \left[\int_0^t e^{p_n(t-\tau)}u(\tau)\mathrm{d}\tau\right] + d\cdot u(t) - \sum_n \left[\int_0^t e^{p_n(t-\tau)}y(\tau)\mathrm{d}\tau\right] - d'\cdot y(t) \tag{2.16}$$

Where the convolutions may be handled either by the method of recursive convolutions or auxiliary integration. The time domain method assumes the input, $u(t)$, such that the only unknowns are again the poles and residues of the modeled system. If the auxiliary integration method is carried out using a standard integration method such as trapezoidal integration, the least squares problem is then defined with a new time domain basis:

$$A_{0,:} = \left[\begin{array}{ccccccc} \mathcal{L}\{\frac{U(s)}{s-p_0}\}^{-1} & \ldots - \mathcal{L}\{\frac{U(s)}{s-p_N}\}^{-1} & u(t) & -\mathcal{L}\{\frac{Y(s)}{s-p_0}\}^{-1} & \ldots - \mathcal{L}\{\frac{Y(s)}{s-p_N}\}^{-1} & -y(t) \end{array}\right]$$

$$x = \left[\begin{array}{c} r_0 \\ \ldots r_N \\ d \\ r'_0 \\ \ldots r'_N \\ d' \end{array}\right]$$

$$A_{0,:} \cdot x = \underbrace{y(t_0)}_{b}$$

The bases of Eq. (2.13) and Eq. (2.16) are an improvement in terms of its conditioning when compared to the monomial basis which is typically used for rational polynomial fitting. The Vandermonde matrix that is formed from a monomial basis is notoriously ill-conditioned since polynomials have abscissas with increasing exponentiation. Eventually the difference between the lower order terms and higher order terms becomes so wide that truncation errors begin to occur. This can even happen for lower order systems that are evaluated over a wide range of abscissa values (wide frequency range in this case). The rational bases of Eq. (2.13) and Eq. (2.16) fortunately only have first order poles but severe ill-conditioning can still occur when these poles are nearly identical (and hence some columns of A approach linear dependence) and again when the abscissa spans large ranges. Orthonormal Vector Fitting improves the conditioning of A by orthogonalizing the columns of A[8]. Orthogonalizing the columns of A can be done using an off-the-shelf QR factorization method such that $A = QR$ (where Q is MxM and is orthogonal and R is MXN and is upper-triangular). The normal equations of the least squares solution then reduce to:

$$x = (A^T A)^{-1} A^T = (R^T Q^T Q R)^{-1} R^T Q^T b = R^{-1} Q^T b \tag{2.17}$$

The least squares process outlined in this section applies to solving the actual fitted residues, which comprise the first N columns of A, as well the amount to relocate the poles by, r'_n, which are the last N columns of A (before the constant term). This is why VF is a two stage linear least squares problem. The first stage is only concerned with the first N columns (plus the constant) and the second stage is only concerned with the latter.

When the constant term, d', cannot be assumed as unity, the method calls for Relaxed Vector Fitting (RVF) whereas if the constant is known (for instance if it is fundamentally a universal constant) the method is known to be the classic VF method[9]. This is because in its original publication, d was assumed to be unity for all cases. In section, the least squares solution of VF was treated in its most general case (where no assumptions are made about the constant terms). There are however additional constraints which must be placed on the least squares fit if d' is not unity.

2.4 Relaxed and Non-Relaxed Pole Relocation

While fitting a rational polynomial approximation in a least squares problem, the L2-norm error to minimize is $|Ax - b|_2^2 = |\tilde{H} - H\sigma|_2^2$, where $\tilde{H} = Ax$ are columns corresponding to the sum of terms in (10) for the scalar transfer function approximation to the tabulated column of data, $H = b$, and σ are columns corresponding to the sum of terms for the optimal scaling function defined in (11). If instead the difference error is considered, it is evident that σ biases the result by trying to stay minimally small in magnitude.

$$\begin{aligned} \tilde{H} &\approx H\sigma \\ \tilde{H} - H\sigma &\approx 0 \end{aligned} \tag{2.18}$$

Consider the first stage of the method, pole relocation, where H is not yet solved for. Notice how in Eq. (2.18) if both coefficients of $\tilde{H}$ and σ are small, the absolute difference will be much smaller and thus the LS cost function is minimized. Therefore, when the initial guess is far from the actual location of the system's poles, σ may not necessarily converge since it favors having a smaller magnitude (and large pole updates require larger residue coefficients of σ). This is usually only problematic when Eq. (2.18) cannot be satisfied exactly either due to the observation not being

spanned by the VF basis or by AWGN corrupting the observation. In these cases, it may be more advantageous to reduce the magnitude of σ rather than relocate to a correct pole location.

Relaxed VF (RVF) reduces the magnitude of pole relocation bias by allowing the constant term of σ to take on any real value to compensate for large residue coefficients. The constant term of σ was previously described in the frequency domain by Eq. (2.13) as $1 + d'$. For non-relaxed VF, $d' = 0$ and so σ must approach unity at infinite frequency. In this case, the least squares problem of Eq. (2.14) has one less column.

$$\underbrace{\begin{bmatrix} \frac{1}{s-p_0} & \dots \frac{1}{s-p_n} & 1 & -\frac{H(s)}{s-p_0} & \dots -\frac{H(s)}{s-p_N} \end{bmatrix}}_{A_{0,:}} \cdot \underbrace{\begin{bmatrix} r_0 \\ \dots r_N \\ d \\ r'_0 \\ \dots r'_N \end{bmatrix}}_{x} = \underbrace{H(s)}_{b}$$

By constraining σ to have a constant coefficient of unity, σ has less incentive to relocate poles further. By implementing the RVF solution, the coefficient may be much smaller than unity by having a negative d'. However, since d' is no longer constrained, it is also possible for to be zero (which is a trivial solution for minimizing the LS cost). To constrain σ to be a non-zero value, an additional row is added to Eq. (2.14) asserting this.

$$A = \begin{pmatrix} \frac{1}{s_0-p_0} & \dots & \frac{1}{s_0-p_N} & 1 & -\frac{H(s_0)}{s_0-p_0} & \dots & -\frac{H(s_0)}{s_0-p_N} & -H(s_0) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{s_M-p_0} & \dots & \frac{1}{s_M-p_N} & 1 & -\frac{H(s_M)}{s_M-p_0} & \dots & -\frac{H(s_M)}{s_M-p_N} & -H(s_M) \\ 0 & \dots & 0 & 0 & Re\left\{\sum_{m=0}^{M} \frac{1}{s_m-p_0}\right\} & \dots & Re\left\{\sum_{m=0}^{M} \frac{1}{s_m-p_N}\right\} & \sum_{m=0}^{M} 1 \end{pmatrix}$$

$$A \quad \cdot \quad \underbrace{\begin{bmatrix} r_0 & \dots r_N & d & r'_0 & \dots r'_N & d' \end{bmatrix}^T}_{x} \qquad = \qquad \underbrace{\begin{bmatrix} H(s_0) \\ \vdots \\ H(s_N) \\ M \end{bmatrix}}_{b} \tag{2.19}$$

values are taken assuming there should always be a finite real portion to σ while the imaginary portion may be zero (such as in the time domain when measurements are purely real). The last row of A in RVF is then weighted by the size of H and the size of the observed frequency domain, $RVFweight = \frac{|w \cdot H|_2}{M}$.

This weighting scheme incites how small σ should be relative to the size of H. When H has an overall smaller magnitude this relaxes σ to have a smaller magnitude as well. To prevent outliers such as H having several high-Q resonances, the weight is divided by the number of points which essentially averages the sums in the last row of Eq. (2.19).

When implementing the normal equations of Eq. (2.17), the real and imaginary portions of A and b are concatenated vertically (stacked) such that the x vector satisfies both real and imaginary portions simultaneously. When this is done for RVF, the rows have a linear dependence since the last row is the real portion of the sum of all the columns of . Therefore, after a basis is formed for the row space of A using QR decomposition, the linearly dependent rows are removed and only the last row of Q remains. Because the solution for the poles only concerns the terms, this further reduces Eq. (2.17) to:

$$x = R^{-1}_{N+1,N+1} Q^T_{M+1,N+1:} b$$

Where $M+1$ signifies the last row of the matrix, and $N+1$ signifies the columns

to the right of the N^{th}order approximation plus its constant term (assuming the constant is being solved for as well).

RVF has the advantage of being able to locate the poles of the system more accurately when noise is present than the original VF algorithm. This is simply because the cost function associated with large relocation (and hence large residues of σ) may be greater than if were to only perform a smaller relocation. Because most end-users prefer to not have to provide a good initial guess the of the system poles, it is usually customary to always implement RVF. The only disadvantage of RVF is that an additional unknown must be solved for the constant term of σ (which is almost computationally negligible).

2.5 Summary of Vector Fitting Algorithm

VF and its extended implementations (time domain, relaxed vector fitting, etc.) can be generalized into the following paraphrased steps:

1. The poles of the system are solved for, given an initial guess and initial number of poles.
   (a) The r'_n and d' terms of σ are solved for, neglecting r_n and d terms
   (b) The SSR of Eq. (2.12) is used to relocate the current best guess of the poles proportional to the just-solved residues of σ'.
   (c) Unstable (positive) poles are flipped (negated) so as to force them to be stable.
2. The residues of the system are solved for given the latest best guess for the system's poles.
   (a) The new poles update the basis functions in the columns of A.

(b) The r_n and d terms are solved for, neglecting r'_n and d' terms.

3. The convergence criteria are checked and the algorithm exits if the fit is suitable. Otherwise the process restarts at step 1 using the latest values of poles and residues for the approximation.

Chapter 3

ESTIMATION OF MULTI-DEBYE PERMITTIVITY

3.1 Summary

The Vector Fitting (VF) method [1] is applied to obtain a deterministic estimate for the permittivity of a Debye material when measured in the presence of interferences caused by test setup imperfections. Measurements are conducted using an admittance tunnel [10] with a resonant iris that is shown to be a significant source of interference. Expressing the Debye frequency dependence of the measured permittivity as an analytic RC circuit model [11], all RLC Lorentz resonances are detected as interference and are subtracted out of the model. A full-wave simulation is performed to measure the admittance of the iris structure in an ideal admittance tunnel, isolated from all other sources of interference. The simulated interference caused by the iris is verified in actual hardware measurements. VF is shown to obtain accurate estimates of the true sample permittivity, when in the presence of these interferences, as long as the signal-to-interference-ratio (SIR) is sufficiently high such that the Debye relaxations are observable in the data.

3.2 Introduction

The admittance tunnel [10] is a free-space test setup that is advantageous for performing non-destructive material permittivity and permeability measurements. To perform a correct S_{21} transmission measurement, the incident fields should not leak around or bypass the material sample under test. To prevent this from happening, the material is placed adjacent to a narrow aperture or *iris* (as shown in Figure 3.1),

which ensures that fields are confined to pass through the material.

The iris is cut out of a large ground plane that also provides a convenient platform for the material sample to rest on. The shape of the iris, such as the diamond shape used in this work, is optimized to reduce scattering [12].

The irises used in admittance tunnel measurements are resonant structures and are therefore frequency selective [13]. At low frequency (longer wavelengths), small irises are good reflectors of the incident energy since there is strong diffraction from its Perfect Electric Conductor (PEC) edges. This phenomenon is also known as "low frequency choking", which is further discussed in Section 3.3.2.

To reduce sources of interference in the measurement, usually the imperfections are included in the calibration. This means that a unity transmission through the open iris in air is first measured and then all further material measurements are conducted with the iris covered by the material. The final result is reported as the material measurement normalized by the unity transmission measurement in an attempt to only realize relative changes in S_{21}.

No matter how precisely the calibration is performed, the material under test may still tune the resonances of the iris to a shifted frequency, unaccounted for by the calibration. This error is especially prone to measurements of low loss, high dielectric materials, as they do not dampen the resonance and constitute a greater tuning capacitance. When the complex poles of the iris are shifted due to the presence of the material, the interference is again observable in the measurement.

A more rigorous approach to removing the interference is to detect and subtract the resonant curves in the calibrated measurement data. This can be done accurately by modeling the measured admittance as a sum of Debye relaxation RC circuits and Lorentz tank resonant RLC circuits [11].

$$\epsilon_\infty + \sum_n \frac{\triangle\epsilon_n}{1 + j\frac{\omega}{\omega r_n}} + \sum_m \frac{\triangle\epsilon_m}{1 + j\frac{\omega}{\omega r_m}\frac{1}{Q} - \left(\frac{\omega}{\omega r_m}\right)^2} \tag{3.1}$$

Where for either an *nth* Debye or *mth* Lorentz term in Eq. (3.1), $\triangle\epsilon$ is the material constituent, ωr is the relaxation radian frequency, Q is the quality factor of the Lorentz resonance, and ϵ_∞ is the high frequency asymptote for the permittivity. The equivalent circuit elements are given directly by the Debye relaxations and Lorentz resonances:

$$\omega r_n = \frac{1}{R_n \epsilon_0 \triangle\epsilon_n} \tag{3.2a}$$

$$\omega r_m = \frac{1}{\sqrt{L_m C_m}} \tag{3.2b}$$

$$Q = \frac{\omega r_m L_m}{R_m} \tag{3.2c}$$

The Debye and Lorentz terms in Eq. (3.1) form a complete basis for the sample admittance in an ideal measurement, and their respective coefficients in Eq. (3.2a) through Eq. (3.2c) may be estimated from a least squares solution using Vector Fitting (VF) [1]. In Section 3.5 it is shown that the Lorentz resonances of the iris interference may be identified by VF and subtracted from the circuit model, yielding a deterministic estimate of the Debye material permittivity. This process is referred to as complex pole elimination.

Alternative procedures for modeling Debye materials with VF have been proposed in [14, 15**? ?**], but enforcement of passivity is at best executed by removing or replac-

ing terms that are in violation, rather than directly constraining the VF objective function to have positive linear terms. In this work, only passive circuit elements are estimated for modeling Debye materials by constraining the least squares solution of VF to have residues with positive real components, as is done for the general curve-fitting procedure in [5].

To ensure that VF is removing only spurious poles due to the iris, an ideal admittance tunnel is simulated and the admittance of the iris is calculated separately from all other hindrances on the measurement in Section 3.4.2. The location of the spurious poles found in the simulated iris admittance are used as a reference that justify the removal of complex poles from all future measurements when they occur in approximately the same frequency band.

This work provides a procedure for estimating deterministic, analytic Debye functions for the permittivity of a measured, dielectric material. However, if a Lorentz material were to be measured, this procedure may still apply if the complex poles due to the material are not proximate to frequency bands containing the interference. Furthermore, this procedure assumes that, after a proper calibration, all further sources of interference may be modeled as RLC tank circuits with complex poles (that there are no sources of interference which exhibit Debye relaxation).

3.3 Admittance Tunnel Iris Interference

3.3.1 Simulation Model Description

A finite element method (FEM) experiment is designed, using ANSYS High Frequency Structural Simulator (HFSS), to simulate an ideal admittance tunnel with the only source of interference being the iris. This means that without the iris present, a material's permittivity is inverted exactly without any post-processing or correction

of the data. For an ideal test setup, the admittance tunnel is modeled as a lossless vacuum box with Perfectly Matched Layers (PML) on all six sides to absorb scattered fields from the iris and material obstacles.

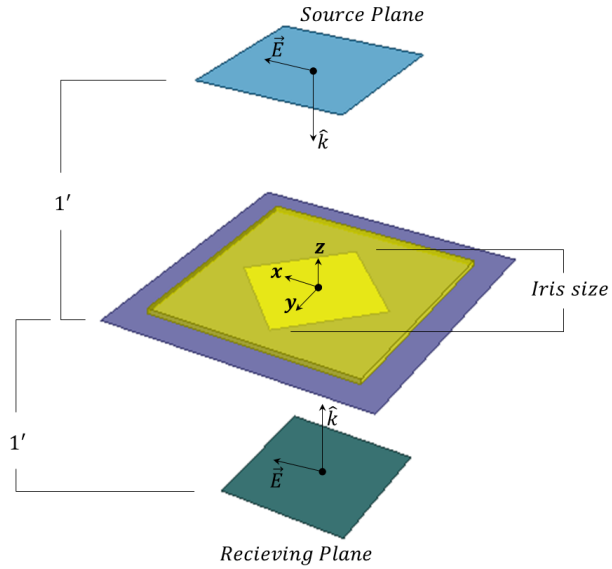

Figure 3.1: Admittance tunnel with material (yellow) resting on top of an iris ground plane (violet) centered at the origin.

The structure is excited by an analytically generated Gaussian beam wave that is injected one foot above the iris in the *source plane* of Figure 3.1, and propagates in the $-\hat{z}$ direction through the iris. The total transmitted fields are measured in the *receiving plane* of Figure 3.1. An S_{21} measurement is then calculated from the measured total fields in the source plane and receiving plane using the reaction theorem (discussed further in Section 3.4.1).

3.3.2 Visualization of Iris Choking

It is well known that an aperture in a metal ground plane that is smaller than approximately one half wavelength is a strongly scattering obstacle, both reducing the transmitted power and introducing diffracted fields into the measurement space. This phenomenon is commonly referred to as "low frequency choking" and is visualized in

Figure 3.2 and Figure 3.3.

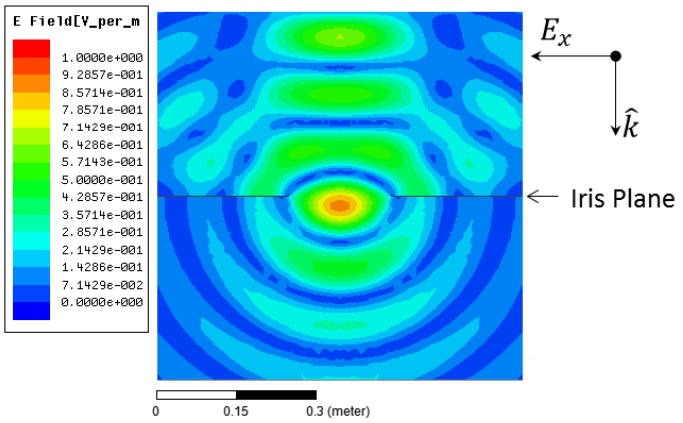

Figure 3.2: $|\overrightarrow{E}|$ fields of 1.5 GHz Gaussian beam wave source in xz plane with 10 inch iris present.

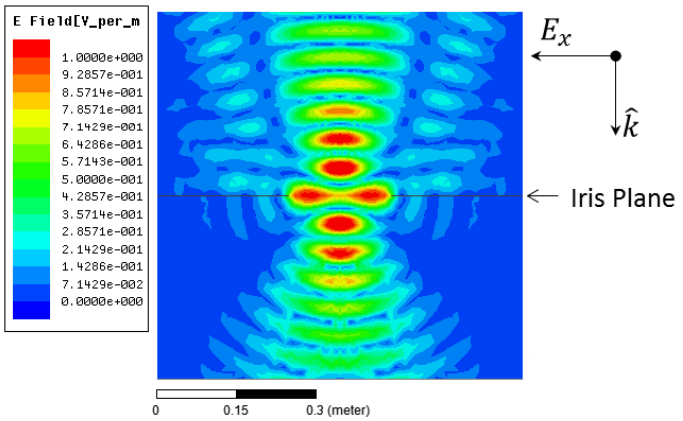

Figure 3.3: $|\overrightarrow{E}|$ fields of 3 GHz Gaussian beam wave source in xz plane with 10 inch iris present.

Since the material is always measured in the presence of an iris, this iris resonance can be misinterpreted by permittivity extraction algorithms to mean that the material under measurement contains Lorentz resonances, instead of just its true multi-Debye behavior. To isolate the permittivity of the material under test from this source of interference, the pole singularities of the iris admittance in the complex frequency domain responsible for these resonances should be located and then removed from all subsequent material models.

3.4 Iris Admittance

The admittance of the iris is calculated from a simulated S_{21} measurement with only the iris present to characterize its resonant behavior. The reaction theorem is applied to calculate S_{21}, which is theoretically based on the equivalent Schelkunoff currents enclosing the source radiator and receiver. This approach is theoretically exact and reduces the simulation time and accuracy by committing actual antenna structures as radiator sources from the model.

3.4.1 S21 Calculation Using the Reaction Theorem

By the Schelkunoff equivalence theorem, a source of radiation can be treated equivalently by a distribution of electric and magnetic currents on a surface enclosing the source of radiation [16]. The reaction theorem then involves the product of the total electric and magnetic fields with their respective Schelkunoff aperture currents, $\vec{K_e}$ and $\vec{K_m}$.

For the case of the admittance tunnel, the receiver aperture is located one foot below the iris plane and thus the received fields are measured throughout an infinite (though numerically truncated) plane one foot below the iris. The electric, K_e, and magnetic, K_m, equivalent aperture currents are calculated from the$-\hat{z}$directed Gaussian source wave with only an $\hat{x}$ component electric field of 1 V/m and $-\hat{y}$ component magnetic field of 1/377 A/m. To account for the z dependent Gaussian taper of the wave's *beam waist*, w_0 (which is the diameter of the spot size), the expression for Schelkunoff currents are also scaled with a Gaussian distribution as shown in Eq. (3.3a) and Eq. (3.3b).

$$\vec{K_e} = \hat{n} \times \hat{H} = -\hat{x}\frac{1}{377}e^{-\left(\frac{\sqrt{x^2+y^2}}{w_0}\right)^2} \tag{3.3a}$$

$$\vec{K_m} = \hat{E} \times \hat{n} = -\hat{y} e^{-\left(\frac{\sqrt{x^2+y^2}}{w_0}\right)^2} \tag{3.3b}$$

The complex S_{21} data for the material sample is reported as the reaction with the material present divided by the reaction of the iris with no material is present. The reaction quantity for both cases is given by Eq. (3.4).

$$\begin{aligned} Reaction &= \iint \vec{E} \cdot \vec{K_e} \ \ dS - \iint \vec{H} \cdot \vec{K_m} \ \ dS \\ &= \iint H_y e^{-\left(\frac{\sqrt{x^2+y^2}}{w_0}\right)^2} dS \\ &\quad - \frac{1}{377} \iint E_x e^{-\left(\frac{\sqrt{x^2+y^2}}{w_0}\right)^2} dS \end{aligned} \tag{3.4}$$

3.4.2 Iris Admittance Calculated from S_{21}

The admittance of the iris is found by relating the S_{21} measurement to the ABCD matrix realization for a shunt obstacle in a transmission line. The impedance and admittance of this topology can be summarized in terms of an ABCD matrix, which can be converted into an S parameter matrix [13], where the iris admittance, Y_{iris}, is then found using Eq. (3.5).

$$S_{21} = 2\frac{2Y_0}{Y_{iris} + 2Y_0}, \qquad Y_{iris} = Y_0 \left(\frac{2}{S_{21}} - 2\right) \tag{3.5}$$

Where Y_0 is the free-space admittance which constitutes the entire medium of the tunnel without the iris present. The HFSS simulation was conducted using typical

iris sizes of 6, 8, 12, and 16 inches. For each iris size, the admittance is plotted in Figure 3.4 and Figure 3.5.

It was stated in the introduction that the iris interferes with the measurement due to the material *tuning* the location of the iris admittance complex poles, and thus deviating from what the calibration accounted for. This can be equivalently stated as the material shifts the frequency response of the iris admittance. Therefore, Figure 3.4 and Figure 3.5 conclude that there is only significant variation in the admittance of small iris sizes and therefore if those frequency responses are shifted by the presence of the material they will most strongly deviate from the calibration. Thus, smaller irises are stronger obstacles for measurements conducted in admittance tunnels.

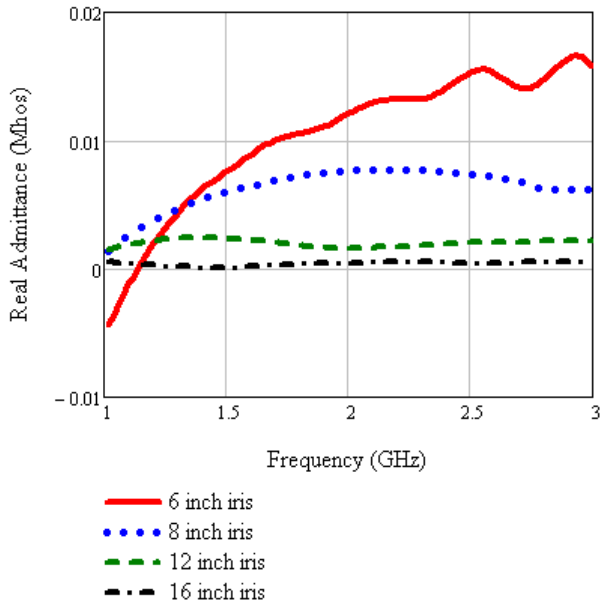

Figure 3.4: Real component of iris admittance with Gaussian beam wave excitation.

The interference due to the iris can be somewhat reduced by truncating the S_{21} measurement in the time domain with *gating*. This simply ignores the delayed reverberations from the resonating iris while still recording the early-time rays which transmit directly through the iris.

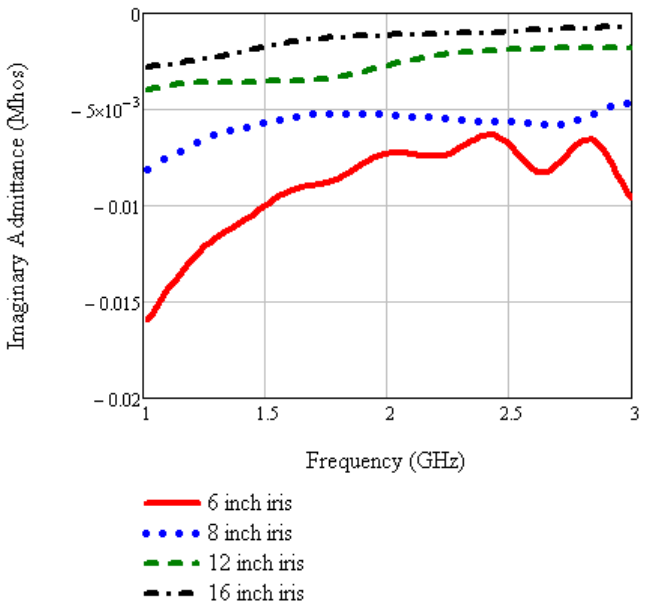

Figure 3.5: Imaginary component of iris admittance with Gaussian beam wave excitation.

3.4.3 Calculation of Permittivity in the Presence of an Iris

The permittivity of a dielectric sample is measured in a simulated admittance tunnel experiment (as described in Section 3.3.1), following the procedure of Section 3.4.1. In order to verify the simulated results, the empirically known permittivity of a Rexolite material ($\epsilon_r \approx 2.45 - 2.55$) is inverted from a physical hardware experiment with an actual admittance tunnel.

The permittivity may be calculated from the S_{21} measurement by applying Davidenko's method [**?**], a complex root solver that implements a differential form of Newton's method. In this approach, the root that is being searched for is the zero produced when subtracting the measured S_{21} from the theoretical S_{21} of a plane wave through a finite thickness slab with intrinsic impedance given by the iteratively guessed permittivity and permeability [17].

As expected, the resonant behavior of the iris is observable in both the Rexolite

material simulations and measurements shown in Figure 3.6 and HFSS(3.7). Due to noise variance and late-time echoes from the surrounding test facility appearing in the measurement, the iris signature is not known deterministically and this is evident when the ungated result matches the simulation better than the gated result.

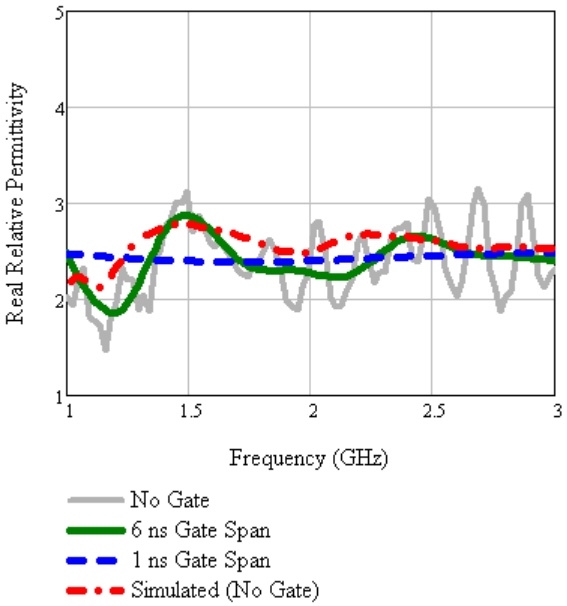

Figure 3.6: Real relative permittivity measurement of Rexolite sample with 16 inch iris, 6 ns gate and ungated measurement.

Typical iris sizes have been shown to significantly interfere with the measurement below 3 GHz. There is thus a trade-off between choosing a conveniently small iris for a given finite sample size and reducing iris interference. The level of interference is best described in terms of a signal-to-interference ratio (SIR). In this context, the frequency response of the complex poles are the interferences that corrupt the true frequency response of the Debye permittivity, or *signal*.

3.5 Results

The interference of the iris may be removed from the final permittivity estimate by locating all complex poles in the measurement and subtracting them from the

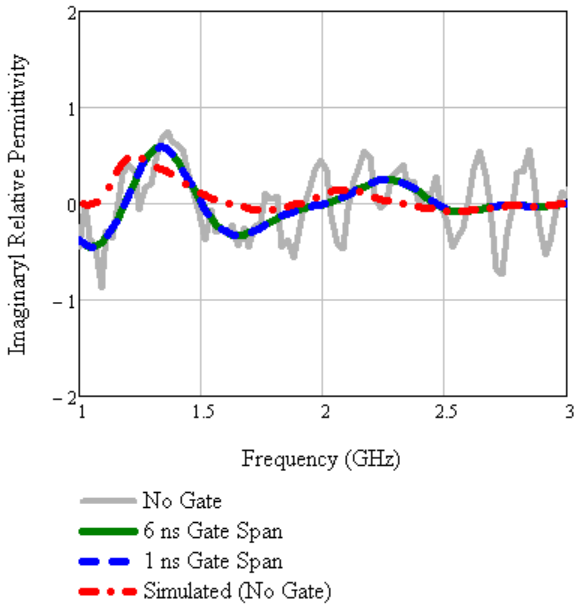

Figure 3.7: Imaginary relative permittivity measurement of Rexolite sample with 16 inch iris, 6 ns gate and ungated measurement.

analytic model produced by VF. The remaining information in the measurement after the subtraction consists of only a constant term and sum of real Debye terms. To correctly apply VF, three major guidelines should be followed as summarized below:

1. The SIR must be high enough such that the actual Debye relaxations are observable in the data.

2. Undesired resonances in the inverted permittivity must be proven to be spurious poles of the test setup before justifying their removal.

3. Spurious poles of the test setup should not exist near the same resonant frequencies of a Lorentz material under test.

It is emphasized that the detected spurious poles of the test setup may only be used as a reference for elimination. This is because the exact location of the complex poles may be easily tuned by the presence of the dielectric material under test as was shown in Section 3.4.2.

3.5.1 *Enforcement of a Passive and Stable Model*

When estimating the permittivity of the sample under test using VF, the poles and residues are solved as the coefficients of Eq. (3.1). The result is represented in the equivalent pole-residue form as shown in Eq. (3.6).

$$\epsilon_r(s) = \epsilon_\infty + \sum_n \frac{r_n}{s - p_n} \tag{3.6}$$

Where $s = \sigma + j\omega$ is the complex radian frequency, and the residues, r_n, and poles (in radians), p_n, are in general complex quantities with associated complex conjugate terms. Pole stability can be enforced, such that p_n are negative, inherently by the VF *pole flipping* approach [1]. However, further constraints must be implemented to guarantee the residue coefficients are positive, thus producing a passive model.

To guarantee passivity, a non-negative least squares method can be used to constrain the purely real residues to be positive. This can be done by replacing the conventional least squares solver used by VF with the freely available LAPACK non-negative least squares implementation. This is sufficient passivity enforcement for this particular application since only the purely real poles are being constrained, and all complex poles are discarded as interference.

If it is desired to also enforce complex poles to be passive, this objective can be more generally stated as a convex optimization problem [5]. Also, non-passive violations may instead be replaced or removed from the model using the approach taken in [**?** **?**]. Finally, there are more rigorous matrix perturbation methods used to enforce passivity such that the matrix of the normal equations is positive semi-definite [6].

Otherwise, an even simpler passivity enforcement approach is to manually increment the order of the VF result to reach the highest possible accuracy until non-

passive or complex poles begin to appear in the model [14, 15]. This approach is most useful for measurements that have a low SIR since in these cases the complex poles are dominant and thus eliminating them will not help resolve the Debye relaxations. In other words, the next best approximation may be to fit the closest Debye curve to the data, without considering the existence of complex poles.

3.5.2 Debye Permittivity Models Generated by Vector Fitting

A Gaussian beam antenna, fed by a Vector Network Analyzer (VNA), is used to radiate the approximately Gaussian wave in an actual hardware admittance tunnel permittivity measurement. The VF estimate of the permittivity for the Rexolite measurements with a 16 inch iris and 1 ns gate are shown in Figure 3.8 and Figure 3.9. Since in this case much higher frequencies are measured that suffer far less interference, the SIR is significantly higher and the result is nearly exact.

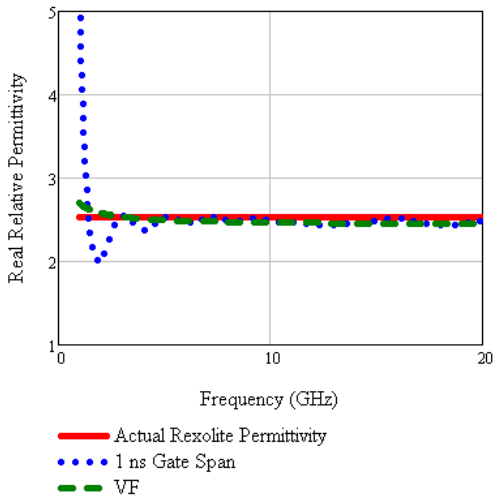

Figure 3.8: Real relative permittivity hardware measurement with 8 inch iris and estimation by VF.

Although no multi-Debye materials with confidently known permittivity were readily available, a simulated measurement of an analytically defined multi-Debye material was conducted. Though the simulated bandwidth is very narrow, using a 16

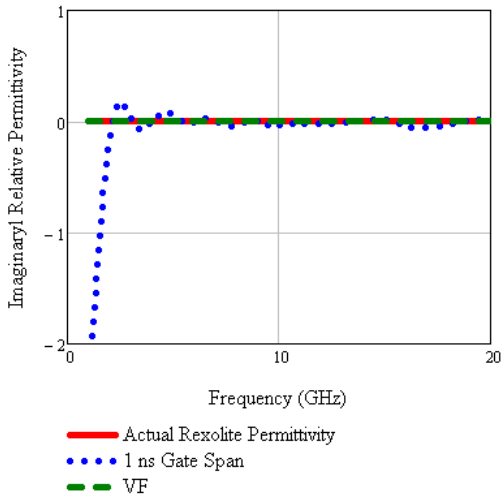

Figure 3.9: Imaginary relative permittivity hardware measurement with 8 inch iris and estimation by VF.

inch iris doesn't incur dominant spurious poles and the SIR is high enough for the true Debye relaxations to be detected by VF, as shown in Figure 3.10 and Figure 3.11.

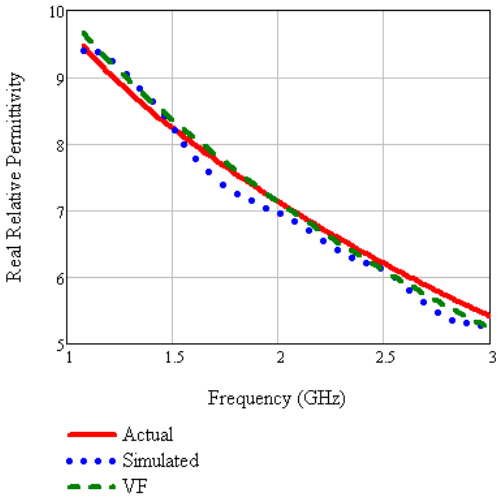

Figure 3.10: Real relative permittivity simulated measurement with 16 inch iris and estimation by VF.

The capability for obtaining deterministic estimates of multi-Debye materials is further exemplified for the case of an absorbing honey-comb sample, as shown in Figure 3.12 and Figure 3.13. Although the material permittivity is not known prior

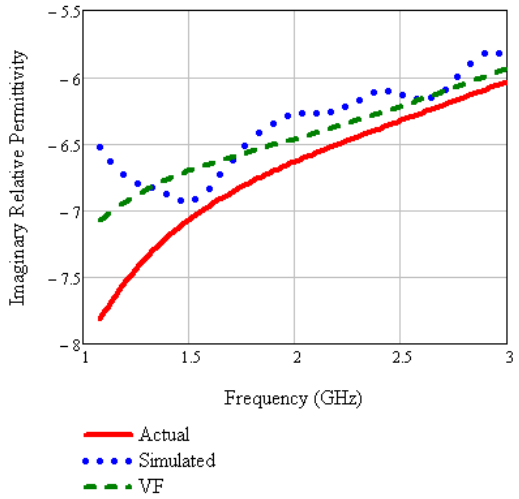

Figure 3.11: Imaginary relative permittivity simulated measurement with 16 inch iris and estimation by VF.

to measurement, the goodness of fit can only be judged by how well the curve matches the data at high frequencies (where interference is minimal) while still providing a causal estimation of the permittivity. Since this absorbing medium is known not to resonate within microwave frequencies, the complex pole elimination approach may be confidently applied to detect only the Debye relaxations.

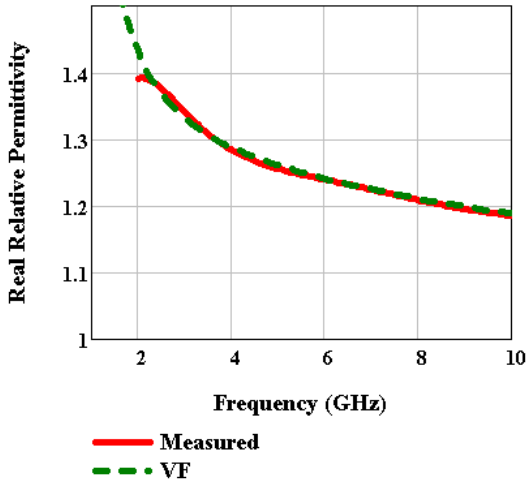

Figure 3.12: Real relative permittivity of honeycomb absorber using a 6" iris and 500 ps gate span.

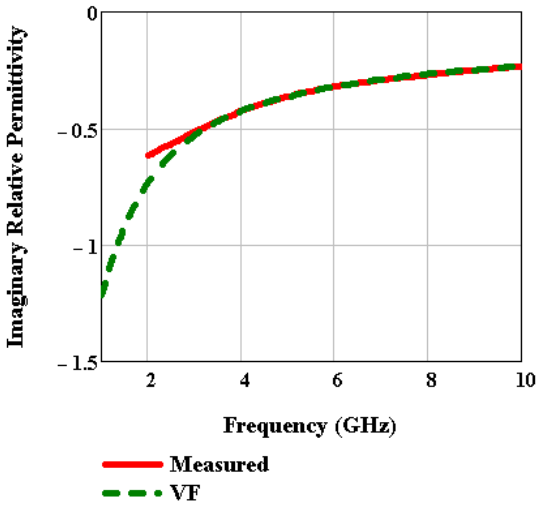

Figure 3.13: Imaginary relative permittivity of honeycomb absorber using a 6" iris and 500 ps gate span.

It has been shown that VF can be applied to estimate a material's permittivity as an analytic, passive, and stable circuit model. By following the guidelines listed at the top of this section, the resulting model can be sufficiently accurate if not exact. In scenarios involving measurements with high SIR and narrow bandwidth, the VF result may still be of use as a passive, circuit-synthesis tool.

3.6 Conclusions

The permittivity of a multi-Debye material has been estimated from measurements conducted with test setup imperfections that cause spurious complex poles to arise in the data. The VF method is shown to accurately locate the spurious poles such that they may be eliminated from the resulting model. This method is accurate when the SIR is high enough such that the Debye relaxations are clearly evident in the data.

An admittance tunnel test setup with resonant iris has been characterized and the spurious complex poles were removed in both simulation and measurement. The simulation and hardware results were nearly equivalent, suggesting further that the

pole elimination procedure implemented in this work is a practical method for material measurement data post-processing.

Chapter 4

ESTIMATION OF DIELECTRIC MATERIAL ADMITTANCE

4.1 Summary

The resonant cavity perturbation method, developed by Waldron [18, 19], is shown to produce errors in excess of 20% in both the real and imaginary components of the permeability for thin, rectangular samples when the demagnetization factor is calculated using the approach in [20]. An alternative approach is recommended to express the permeability of the sample as a function of the effective (or apparent) permeability, which is measured by the Waldron perturbation equations. This relationship between the effective and sample permeability is formulated as a Morphology function [21], with Debye poles that are estimated using the Vector Fitting (VF) method [1]. It is shown that any value of sample permeability can be inverted from a single Morphology function, incurring at most 10% error in real and imaginary components.

4.2 Introduction

To measure the permeability of thin substrate materials at Radio Frequencies (RF), transmission line methods are often employed since they conveniently house the sample under test (SUT) between their conducting terminals (where the fields are highly concentrated), allowing sensitive measurements to be made. Furthermore, since Transverse Electromagnetic (TEM) transmission line modes can exist for any operating wavelength, electrically small SUTs may be measured.

The shorted stripline resonant cavity is one such transmission line test setup that can individually measure either the permittivity or permeability of a thin material

(thin along the longitudinal direction of the strip). As shown in Figure 4.1, it is a two port test device that supports a quasi-static TEM mode excited by the mutually coupled coax feed loops. The mode structure is illustrated in Figure 4.2.

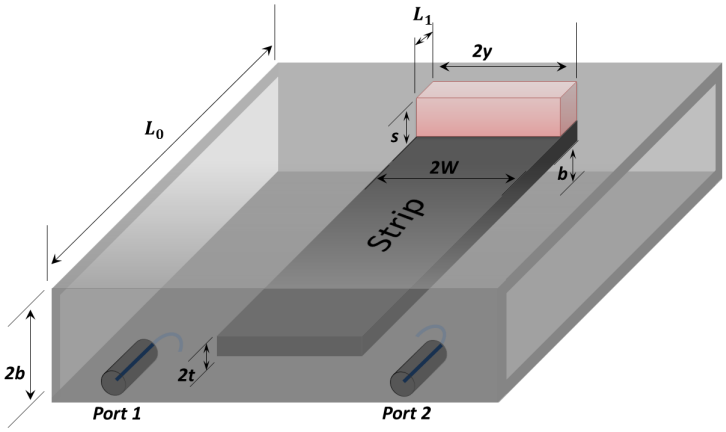

Figure 4.1: Dimensions of the stripline cavity resonator and SUT (red). The strip makes contact on the front wall (adjacent to the ports) and back wall (adjacent to the SUT).

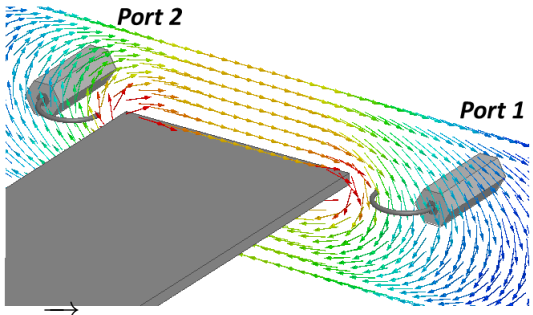

Figure 4.2: Transverse $\overrightarrow{H_0}$ fields that wrap around the rectangular strip and are excited by the two coax loops of Port 1 and 2. The rest of the walls of the cavity are transparent. The $\overrightarrow{H_0}$ field has amplitude of 1 A/m where colored red and 0 where colored blue.

In this work the SUT is measured while placed flat against the back wall, where, due to the short circuit, the interrogating magnetic field, $\overrightarrow{H_0}$, will maximally polarize the material. Furthermore, the electric field is zero such that a perturbation in the cavity is due solely to the permeability of the SUT, μ_{SUT}.

S_{21} is measured both with and without the SUT present to deduce relative changes in the cavity's resonance due to the SUT. Following Waldron's procedure in [19], the

relative shift in resonant frequency and decrease in quality factor ("Q") of the cavity are approximately related to the complex value of μ_{SUT}.

In the derivation of Waldron's perturbation equations, it is assumed that uniform lines of magnetic flux exist in the transverse plane throughout the SUT's cross-section ($2y \times s$). Since the fields are only approximately uniform at the center of the strip, this assumption introduces a primary source of error in the measurement. Especially for wider materials (larger y dimension), as shown in Figure 4.3, a significant correction must be applied to the apparent permeability, μ_{app}, to account for the non-uniformity in the $\overrightarrow{B}$ field.

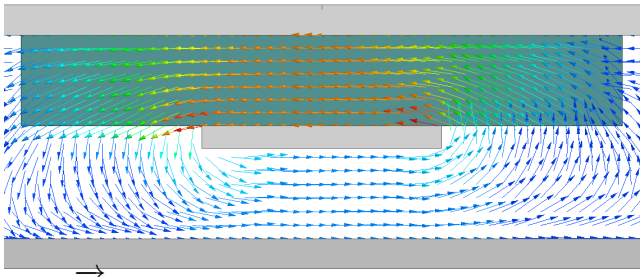

Figure 4.3: Transverse $\overrightarrow{B}$ field plotted in the plane of a $2y = 10$ inch material. The lines of magnetic flux are visibly curvilinear and therefore the assumption of uniform fields is invalid, especially for materials this wide. The $\overrightarrow{B}$ field has amplitude of 1 A/m where colored red and 0 where colored blue.

A second correction is necessary to account for the back magneto-motive force produced by the material since it behaves as a magnetic capacitor. In other words, the interrogating $\overrightarrow{H}$ field induces equivalent magnetic charge distributions where it crosses the finite surfaces of the SUT, and these charges create their own magnetic dipole field that opposes the original field, as shown in Figure 4.4. Although this diminishes μ_{app}, it can be approximately compensated for by scaling μ_{app} with an assumed *demagnetization factor*, N.

The so-called demagnetization factor is only approximate since it arises from the assumption that a uniform depolarizing field, H_D, is induced, which is only exact for

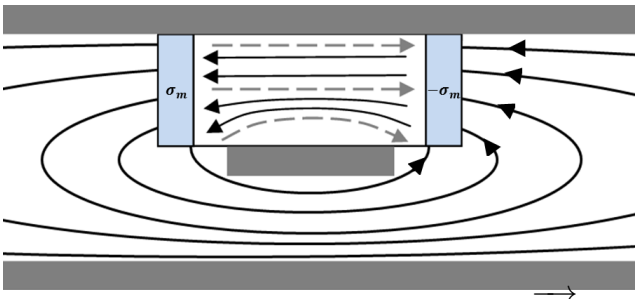

Figure 4.4: Sketch of the internal demagnetizing field, $\overrightarrow{H_D}$ (gray with dashes), in the presence of an applied field, $\overrightarrow{H_0}$ (black), for a SUT with magnetic charge, σ_m, located where the flux paths enter and exit.

ellipsoidal SUT geometries. Under this assumption, the internal magnetic field in the sample is then related to $\overrightarrow{H_0}$, by [**?**]:

$$\overrightarrow{H} = \overrightarrow{H_0} - \overrightarrow{H_D} = \overrightarrow{H_0} - N\overrightarrow{M} \tag{4.1}$$

where $\overrightarrow{M}$ is the magnetization vector and N is the demagnetization factor. Dividing Eq. (4.1) by $\overrightarrow{M}$ yields the susceptibility of the SUT, $\frac{1}{\chi_{SUT}}$, and apparent susceptibility, $\frac{1}{\chi_{app}}$, that is observed in the raw permeability measurement of $\mu_{app} = \chi_{app} + 1$.

$$\frac{1}{\chi_{SUT}} = \frac{1}{\chi_{app}} - N \tag{4.2}$$

As is done in [20, 22], $\mu_{SUT} = \chi_{SUT} + 1$ is then solved for using Eq. (4.3).

$$\mu_{SUT} = \frac{\mu_{app} - 1}{1 - N\left(\mu_{app} - 1\right)} + 1 \tag{4.3}$$

Where the value of N is chosen to best approximate the true depolarization of the material sample in response to the interrogating field.

Since an exact expression for N does not exist for finite rectangular prisms, numerical approaches to modeling the demagnetization can be used [23, 24]. Alternatively, for SUTs placed firmly between the strip and ground plane walls, Aharoni's closed-form solution for the demagnetization of a rectangular prism [25] may apply.

However, N, according to Aharoni's formula, must be further optimized as it was shown in [23, 24] that the demagnetization is also a function of the permeability of the SUT unless it is ellipsoidal in shape.

Aharoni's demagnetization factor, when applied to Eq. (4.3), is a convenient equation that is commonly used in practice [26], and so it will be used as a standard for comparison in this work. For organization, all results utilizing Eq. (4.3) assume the value of N computed using Aharoni's formula and will be labeled by his name in plot legends.

In summary, the two causes of error in μ_{app} that this work proposes improvements for are:

1. The assumption that the externally applied $\overrightarrow{H}$ field has uniform lines of flux passing throughout the entirety of the material.

2. The assumption that Eq. (4.3) exactly inverts any value of μ_{SUT} for an optimal value of N.

Therefore, to provide a general solution for correctly inverting μ_{SUT}, the assumptions listed above mustn't be made. An analytical inversion equation must be formulated that is dependent on both the SUT's permeability and geometry. In other words, the non-linear relationship between the field topology and μ_{SUT}, known as the *Morphology function* [21], must be obtained. The intention of this work is then to find a Morphology function that is parametrized by the SUT dimensions, such that μ_{SUT} of any geometry may be inverted with optimal accuracy.

The approach taken to correctly measure μ_{SUT} involves a two step procedure. First, Waldron's original equations are used to obtain μ_{app} from the relative change in resonant frequency and Q. Second, μ_{SUT} is inverted from μ_{app} using the Morphology function, which relates the change in field morphology when a SUT is introduced to

the unperturbed stripline TEM field topology. This can be done by curve-fitting the morphology function of the stripline cavity resonator using VF.

In Section 4.3 the approach to modeling an ideal measurement of a permeable sample with the stripline resonator is described. For a given measurement of μ_{app}, Section 4.3.2 then shows how a modified version of the Vector Fitting (VF) method may be applied to accurately estimate the corresponding Morphology function. VF results of the Morphology function are summarized in Section 4.4. Section 4.4.3 shows how the maximum range of accurately measurable permeability values can be extended for the stripline resonator by increasing the aspect ratio of the SUT's longitudinal cross-section. Section 4.4.4 shows how the Morphology function can be parametrized by the sample geometry.

4.3 Permeability Measurement and Simulation

A full-wave finite element method simulation of the stripline cavity is designed, using ANSYS High Frequency Structural Simulator (HFSS). Simulated measurements of typical rectangular SUT geometries are used to develop the Morphology functions.

As was done in [27], the Morphology function is found by curve-fitting the relation between the known μ_{SUT} (synthesized for the experiment) and the measured μ_{app} of the test setup cross-section that is only partially filled by the sample, as depicted in Figure 4.4 and Figure 4.3.

4.3.1 Simulation Design and Optimization

To ensure the fields are sufficiently isolated from the external environment, the stripline resonator is shielded by a ground plane that can be several feet wide. As such, it is prohibitively large to simulate as a whole and instead should be partitioned into two sections: one that models the coax feeds (denoted "Feed section") of length

L_{01}, and one that models the sample placed against the shorted end of the cavity (denoted "Shorted section") of length L_{02}.

Although resonance conditions require that the two sections be physically connected in order to form a standing wave, the Shorted section can be treated separately as an approximate reactive load that is connected in series with the Feed section equivalent circuit. Then, the original full length of the stripline is maintained since $L_0 = L_{01} + L_{02}$.

By terminating both sections with a wave port, the resulting S parameters of each section allow them to be theoretically reconnected offline to form the full stripline cavity model. This can be performed by first exporting the raw S matrix of both sections from HFSS as touchstone ("SNP") files. Then, they can be cascading together in an equivalent circuit using a circuit simulator, such as Agilent's Advanced Design System (ADS). Figure 4.5 shows how the Feed section and Shorted section are theoretically combined to form the full stripline resonator.

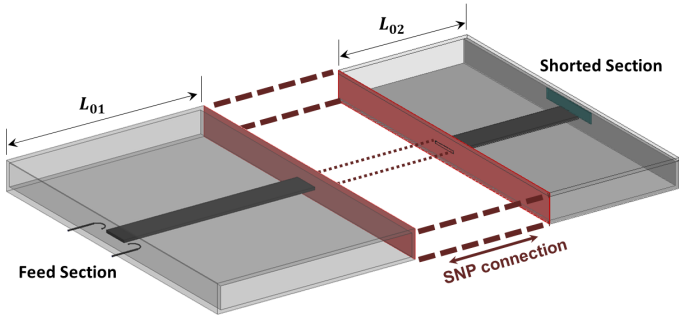

Figure 4.5: "Feed" and "Shorted" section of the numerically partitioned stripline resonator. Both have terminating wave ports (shown as red surfaces) that allow their S matrices to be theoretically cascaded in an equivalent circuit. *SNP* touchstone files are a standard file format that can facilitate this.

The advantage of this method is that all further variations of SUT geometry and permeability are simulated using only the Shorted section model since the Feed section is common to all. This approximation is sufficiently accurate as long as the lengths of each section (L_{01}, L_{02}) are over a quarter of a wavelength to produce the correct

TEM stripline mode.

The simulated material was chosen to have a width, $2y$, ranging from 1–8 inches and length, L_1, ranging from 0.5–10 mm, which are typical sizes used in practice. Smaller sizes are of no concern since the fields would be nearly uniform throughout the SUT for those cases and correction is hardly necessary.

The simulation time can be significantly reduced by identifying the Morphology function from only purely real, positive μ_{SUT} measurements. The Morphology function is then found by curve-fitting the relationship between μ_{SUT} and μ_{app}, which is done using VF. The values of μ_{SUT} are synthesized prior to each experiment so that they are known exactly. The resulting Morphology function is a sum of Debye terms in the w plane:

$$\mu_{app}(w) = 1 + \sum_{d=1}^{D} \frac{A_d}{1 + \frac{w}{jB_d}} \tag{4.4a}$$

$$w = u + jv = \frac{j}{\mu_{SUT} - 1} \tag{4.4b}$$

Since only real values of μ_{SUT} are measured, this means the Morphology function is observed only along the v axis of the complex w domain. However, complex values of μ_{SUT} may still be exactly inverted since the Morphology function is analytic and can be continued throughout the entire complex w domain. Therefore, curve-fitting the Morphology function along the v axis provides a general solution for materials with loss as well.

4.3.2 Vector Fitting the Morphology Function

To curve-fit the Morphology function, a special case of VF is implemented whereby the purely imaginary abscissa, $s = j\omega$, of the Frequency Domain Vector Fitting

(FDVF) basis is replaced by $v = \frac{j}{\mu_{SUT}-1}$. Since there is a cancellation of j terms in the denominator, the Morphology function is purely real-valued along the v axis, which differs from the standard implementation of the FDVF basis.

$$\mu_{app}(v) - 1 = \sum_{d=1}^{D} \frac{A_d}{1 + \frac{jv}{jB_n}} = \sum_{d=1}^{D} \frac{A_d B_d}{B_d + v} \tag{4.5}$$

Where the quantity, $A_d B_d$, represents the residue of the pole, B_d, for a D^{th} order system. Aside from the modified FDVF basis, the exact same VF algorithm is implemented as originally described in [1]. As demonstrated in Section 4.4.4, this modified basis is not only applicable to fitting a Morphology function, but also to curve-fitting the relationship of $A_d B_d$ and B_d versus sample geometry. This relationship can be used to analytically predict the poles and residues for a Morphology function given the SUT dimensions.

For optimized efficiency, only eight variations of μ_{SUT} are simulated, which is enough known quantities to ensure the least squares formulation in VF remains over-determined (the Morphology VF basis can have at most $2 \cdot D + 1$ unknowns). However, this also means any error in the HFSS simulation will easily propagate into the fitted Morphology function since the continuous v axis has been sparsely sampled (these errors are discussed in Section 4.4.2). Still, VF is advantageous for this application since its pole-residue basis exactly spans the Morphology function (when there is no noise or interference present).

4.4 Results

VF is applied to identify the Morphology function of Eq. (4.5). As discussed in Section 4.3.2, the fit is performed using only positive, real values of μ_{SUT}. The measured value of μ_{app} is reduced from the simulated S_{21} measurements of μ_{SUT} using Waldron's equations [19] without any correction applied for demagnetizing effects

(since the Morphology function is already intended to serve this need).

For better efficiency, only eight variations were performed for μ_{SUT} =4, 10, 20, 40, 80, 320, 1000, 3000. For comparison, an approximate solution for μ_{app} is calculated (the dashed gray curves of Figure 4.6 through Figure 4.9) by inverting Eq. (4.3).

$$\mu_{app} \approxeq \frac{\mu_{SUT} - 1}{(p = 1) + N\,(\mu_{SUT} - 1)} + 1 \tag{4.6}$$

By inspection, the partial fraction term in Eq. (4.6) is a first order Debye expression, which is evident after dividing the numerator and denominator of Eq. (4.6) by $\mu_{SUT} - 1$. Therefore, p represents a pole with residue, $\mu_{SUT} - 1$. Then Eq. (4.6) is a specific case of Eq. (4.5), with $AB = 1$, $B = N$.

After the Morphology function is fit, it can be analytically continued throughout the w plane by evaluating Eq. (4.4a) for complex values of Eq. (4.4b). The same real values of μ_{SUT} are simulated as before but now the permeability is made complex by adding a magnetic loss tangent of 0.5. For succinctness, the results for lossless samples are not shown since they are redundant. In other words, Figure 4.6 and Figure 4.8 are practically the same results as the case of lossless real permeability.

4.4.1 Analytic Continuation into the Complex w Plane

Figure 4.6 and Figure 4.7 show μ_{app} plotted on the v axis for a material sample with dimensions: $s = 1.6875$", $2y = 2$", $L_1 = 0.2$" (denoted "Sample 1"). The figures also show the Vector-fit result and approximate solution for μ_{app} given by Eq. (4.3). For narrow width (small $2y$) materials, Eq. (4.6) is only approximately correct and N clearly must be corrected by an optimized scalar.

Figure 4.8 and Figure 4.9 shows the Morphology curve for an $s = 1.6875$", $2y = 8$", $L_1 = 0.2$" sample (denoted "Sample 2"). Due to the higher aspect ratio of this particular sample, Eq. (4.6) is not valid for any value of N since the curve does not

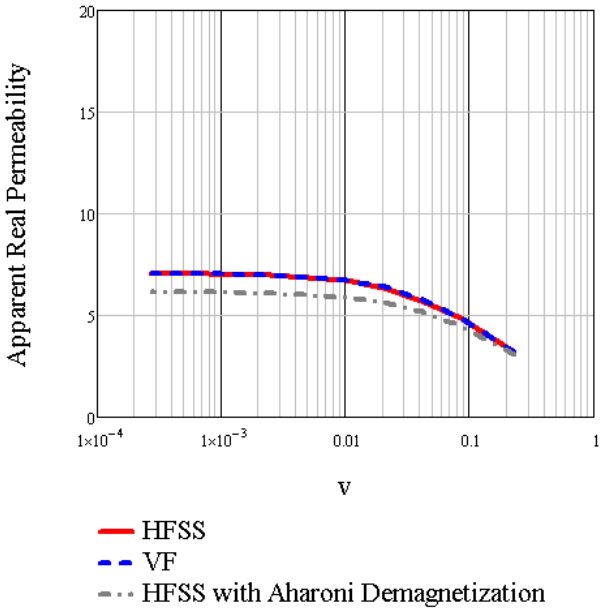

Figure 4.6: Real component of μ_{app} for Sample 1, with dimensions: 1.6875" $\times$ 2" $\times$ 0.2", inside the stripline cross-section plotted along the imaginary $v = 1/(\mu_{SUT} - 1)$ axis.

even form the correct shape.

4.4.2 Inverting μ_{SUT} From the Morphology Function

Any complex value of μ_{SUT} may then be obtained for any measured μ_{app} using the general case of Eq. (4.6) for a first order Morphology function shown in Eq. (4.7).

$$\mu_{SUT} = \frac{\mu_{app} - 1}{A_d B_d - B_d \cdot (\mu_{app} - 1)} \tag{4.7}$$

Figure 4.10 and Figure 4.11 show the calculated μ_{SUT}, for Sample 2, plotted against the actual value known prior to the simulation. In all cases, near equivalence is reached with the true value of μ_{SUT} as further evidenced by the percentage error plots for each measurement in Figure 4.12.

Formulating the inversion of μ_{SUT} as a Morphology function overcomes the limited range of material size that conventional methods using Eq. (4.3) are restricted

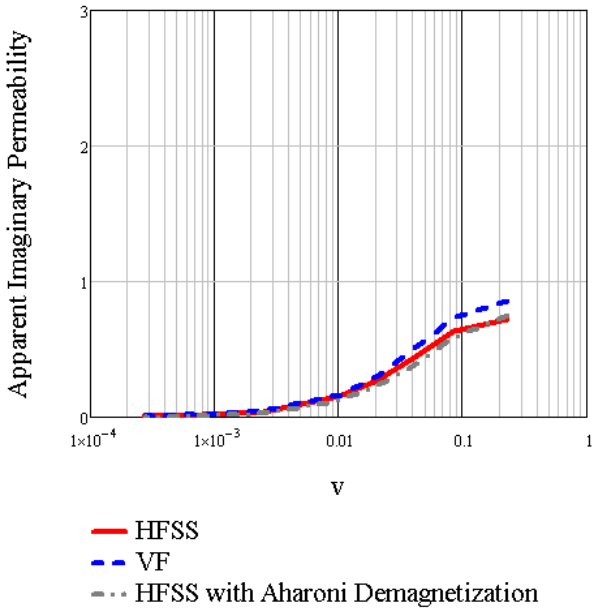

Figure 4.7: Imaginary component of μ_{app} for Sample 1, with dimensions: 1.6875" $\times$ 2" $\times$ 0.2", inside the stripline cross-section plotted along the imaginary $v = 1/(\mu_{SUT} - 1)$ axis.

to. Although N could be further optimized to match the apparent permeability in Figure 4.6 and Figure 4.7 (scaling N by 1.65 seems optimal), the same compensation cannot improve the approximate result of Figure 4.8 and Figure 4.9 since the assumed pole location ($p = 1$) of Eq. (4.6) is no longer acceptable for the sample's dimensions.

The Morphology function should be exact in theory, but round-off error and non-physical data approximated in the HFSS experiments present noise and interference that biases the VF result. As discussed in Section 4.3.2, since only eight values of μ_{SUT} were simulated, each data point significantly weights the least squares error. If necessary, more values of μ_{SUT} could be simulated to remedy this. Also, to reduce the error in the simulation, the convergence criteria of the HFSS simulations could be further restricted, though at the substantial cost of more computer resources.

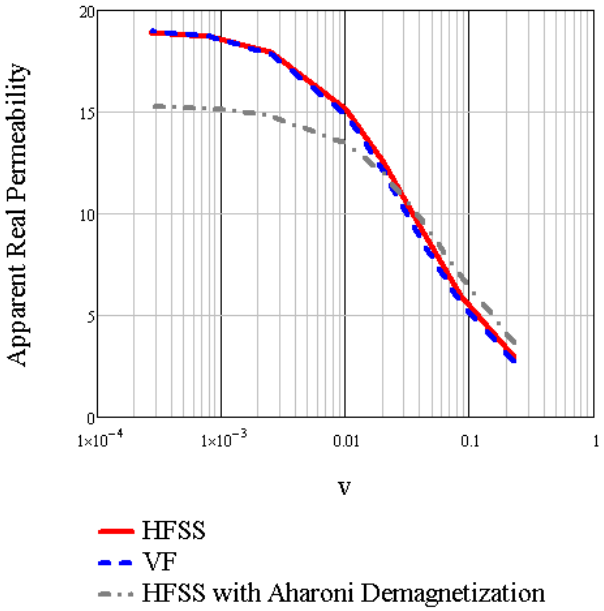

Figure 4.8: Real component of μ_{app} for Sample 2, with dimensions: 1.6875" $\times$ 8" $\times$ 0.2", inside the stripline cross-section plotted along the imaginary $v = 1/(\mu_{SUT} - 1)$ axis.

4.4.3 Morphology Inversion Sensitivity to Sample Polarizability

The inversion error, observed in Figure 4.12 and Figure 4.13, is not only due to simulation noise but also Eq. (4.7) being ill-conditioned for very small and very large μ_{SUT}. This is fundamentally a physical limitation of the material since μ_{app} saturates to a fixed value as μ_{SUT} maximizes or minimizes the polarizability of the sample geometry. Thus, a slight perturbation in small or large values of μ_{SUT} can produce a significantly different result for μ_{app}.

To show this, in Figure 4.14 a marginal 1% deterministic error is added to μ_{app} values that are calculated directly from the Morphology function of three different samples using a sweep of purely real μ_{SUT} values. Then, μ_{SUT} is *re-solved* for by inverting it from the same Morphology functions for each sample but now using the erroneous μ_{app} values.

Due to the sensitivity of the sample polarizability, the error is shown to magnify

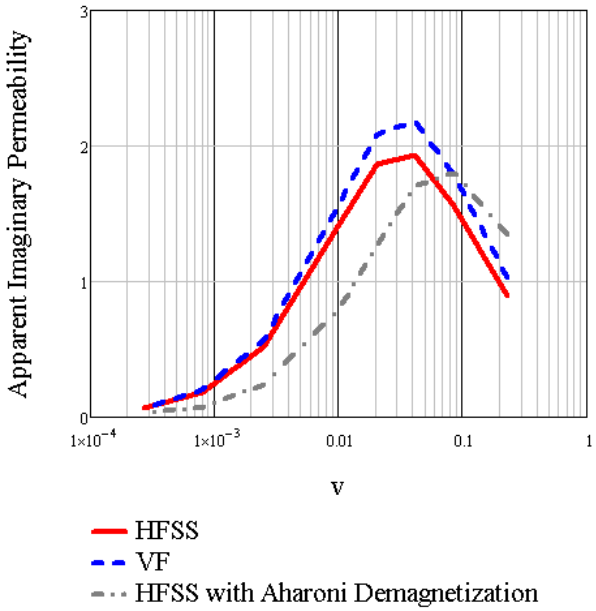

Figure 4.9: Imaginary component of μ_{app} for Sample 2, with dimensions: 1.6875" $\times$ 8" $\times$ 0.2", inside the stripline cross-section plotted along the imaginary $v = 1/(\mu_{SUT} - 1)$.

for low and high values of μ_{SUT} as conjectured. This concludes that the measurable range of μ_{SUT} is inherently limited by the SUT's Aspect Ratio (AR) (determined by the $2y$ and s dimensions).

Manufacturing materials with a higher AR allows them to have larger permeability and still be measured accurately as demonstrated by the green curve of Figure 4.14, which is the error in μ_{SUT} due to a 1% error in μ_{app} for a sample with dimensions $s = 0.442$", $2y = 8$", $L_1 = 0.2$" (Sample 3). This means that the SUT height, s, does not need to make contact with the strip and ground plane, which is contrary to the restrictions in sample dimension that are implied when using the Aharoni demagnetization formula.

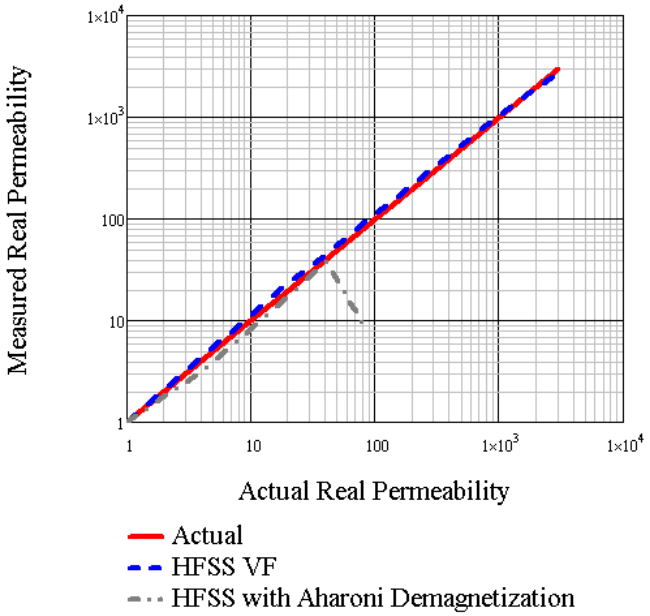

Figure 4.10: Real component of the inverted μ_{SUT} compared to the actual value for Sample 2, with dimensions: 1.6875" × 8" × 0.2", and a loss tangent of 0.5. Note, Eq. (4.6) produces negative results which is why a portion of the gray dashed curve isn't shown.

4.4.4 Parametric Morphology using Vector Fitting

The poles and residues of the first order Morphology function can be parametrized by the sample dimensions by performing a parametric Vector Fitting procedure [28]. This means that two additional equations are formulated that yield the correct pole location and correct residue value for a given material geometry. The simplest way of doing this is to perform a regression on the locations of the morphology poles using a monomial basis. An alternative method that may gain slight insight into the problem, is to curve-fit the relationship between the pole location and material dimension using VF. The benefit to the latter approach is that the pole location appears to trend as an analytic function of the material dimension, which VF is shown to curve-fit nearly exactly.

The modified frequency domain VF basis, as described in Section 4.3.2, is used to

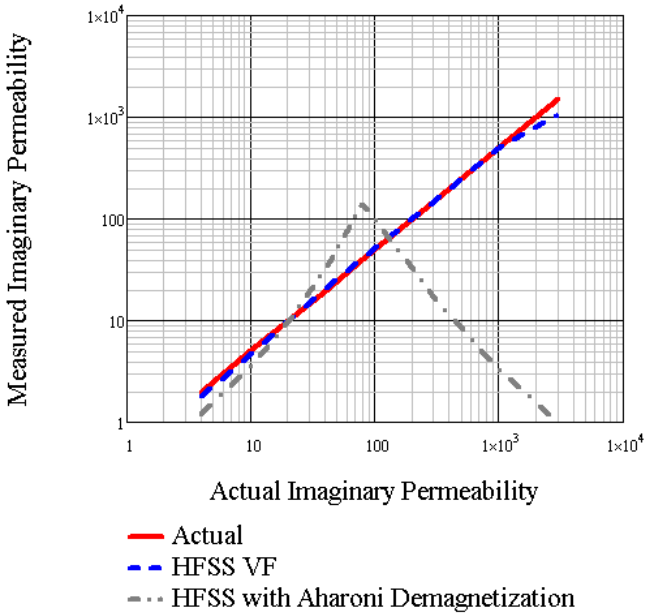

Figure 4.11: Imaginary component of the inverted μ_{SUT} compared to the actual value for Sample 2, with dimensions: 1.6875" × 8" × 0.2", and a loss tangent of 0.5.

curve-fit the pole location since the domain variable (a geometrical length) is purely real. Plotting the location of poles versus material width ($2y$) for several cases of thickness (L_1), the resulting curves are each vector-fit to obtain an analytic partial fraction expression that is precise, as shown in Figure 4.15 and Figure 4.16. The same procedure may be applied for the residues, with results shown in Figure 4.17 and Figure 4.18. In this case, the residue values hardly changed as a function of material thickness and so only one case of thickness is plotted with the VF result.

The fitted curves of Figure 4.15 and Figure 4.17 imply that a unique Morphology function can be determined for any material dimension, simply by interpolating them for a desired material width and thickness. The resulting Morphology function is then constructed by plugging in the interpolated pole and residue values as the coefficients for B and AB in Eq. (4.5).

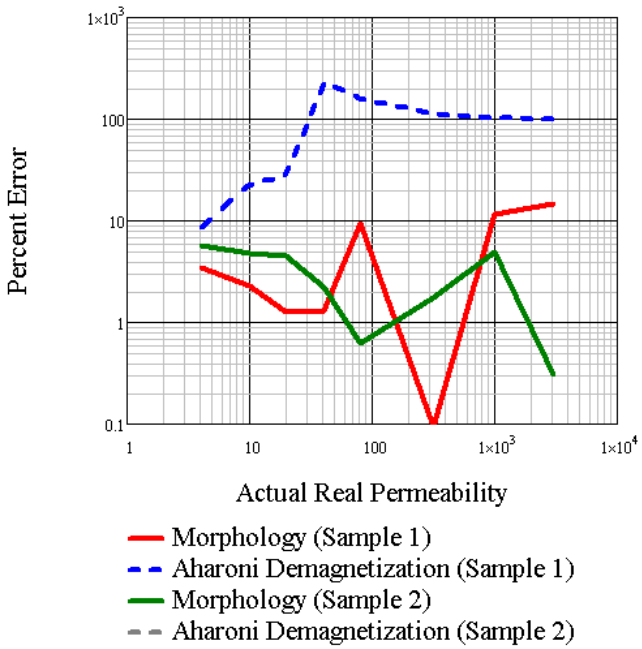

Figure 4.12: Percentage error of the real component of the inverted μ_{SUT} compared to the actual value for Samples 1 and 2, both with loss tangents of 0.5.

4.5 Conclusions

The Morphology for several variations of SUT in a stripline cavity resonator was modeled and verified as an accurate inversion equation for μ_{SUT}. Developing the Morphology function was shown to be numerically efficient, requiring only a few simulated measurements of purely real μ_{SUT} (rather than having to exhaustively model every possible variation of complex permeability).

The application of VF is recommended for detecting the Morphology function since VF uses basis functions that are precisely the same. It is especially advantageous since it obtains a causal estimate of the Morphology function even when sparse simulated measurements are provided for μ_{app}. The method can also be used to curve-fit the relationship of Morphology function pole locations to a sample dimension.

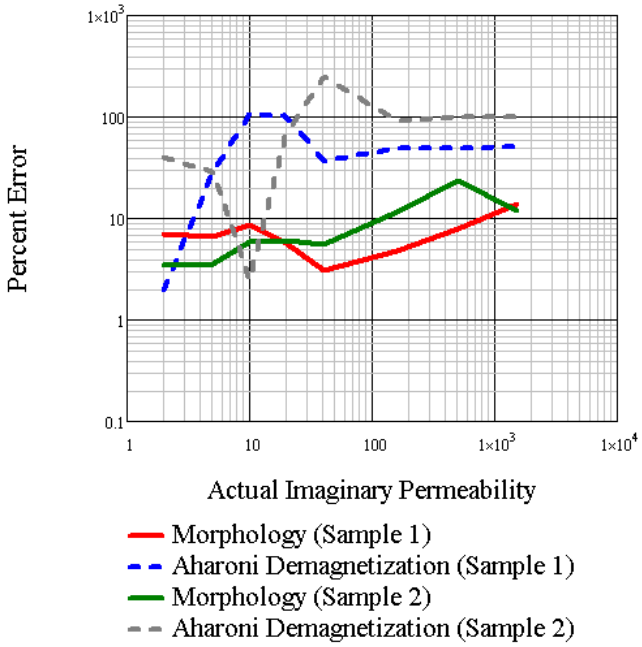

Figure 4.13: Percentage error of the imaginary component of the inverted μ_{SUT} compared to the actual value for Samples 1 and 2, both with loss tangents of 0.5.

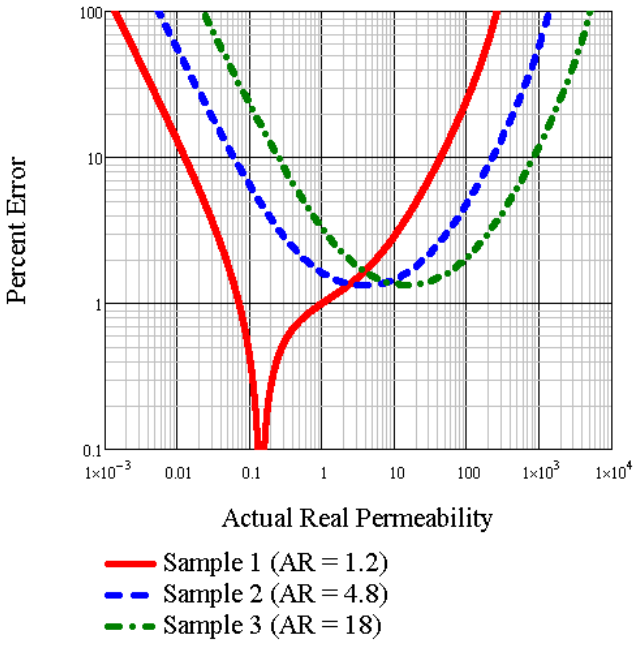

Figure 4.14: Error propagation in purely real μ_{SUT} inverted from a synthesized μ_{app} with fixed error of 1%. μ_{SUT} is inverted for the aspect ratios of Sample 1, 2, and 3 respectively. The error is magnified worst where the Morphology curve has shallow slope (where the polarization saturates).

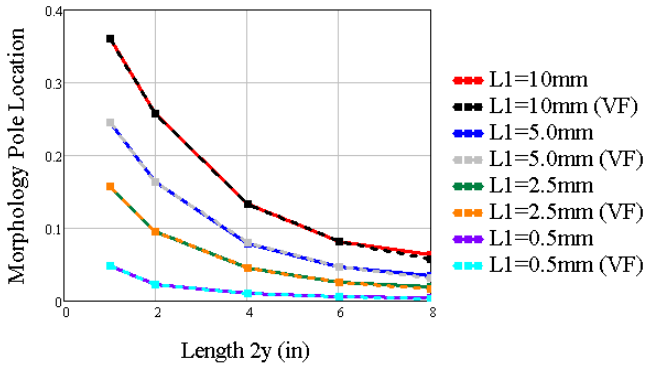

Figure 4.15: Parametric Vector fit of Morphology pole locations as functions of sample width, $2y$, for several cases of sample thickness, L_1.

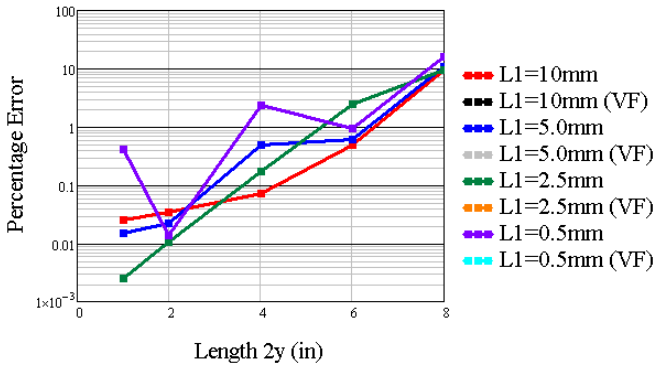

Figure 4.16: Percentage error in parametric VF models of Morphology pole locations as functions of sample width, $2y$, for several cases of sample thickness, L_1.

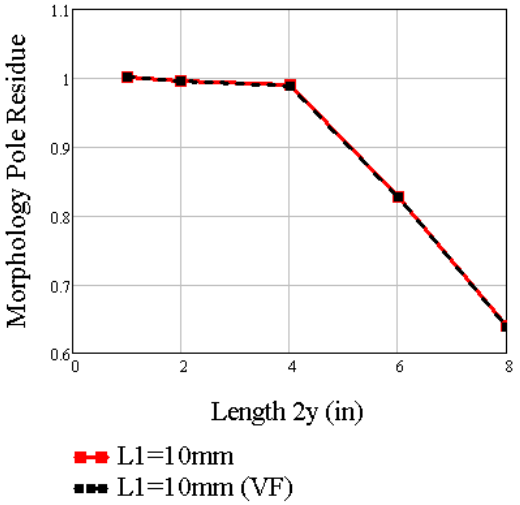

Figure 4.17: Parametric Vector fit of Morphology residue locations as functions of sample width, $2y$. The result is identical for all cases of sample thickness, L_1, and so only one variation is shown.

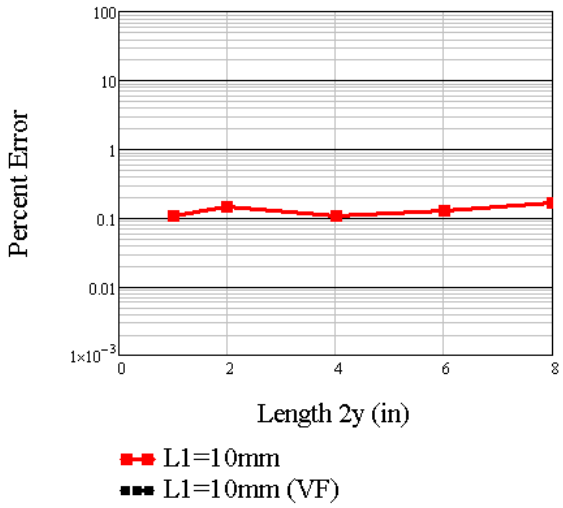

Figure 4.18: Percentage error in parametric VF models of Morphology residue locations as functions of sample width, $2y$, for several cases of sample thickness, L_1.

Chapter 5

ESTIMATION OF TIME DOMAIN SCATTERING HISTORIES

5.1 Time Domain Diffraction

Analytically modeling the time history of diffraction phenomenon in electromagnetics almost exclusively involves asymptotic frequency methods such as the Uniform Theory of Diffraction (UTD). The extraction of diffraction coefficients from canonical scatterers is key to the success of these well known methods. Limited work has been done on the time domain analog of the problem [29, 30, 31, 32].

In particular, Makhoul provides in his work analytical equations for the time history of an infinite 2D wedge in the time domain [29]. Makhoul's equations have been selected in this work to calculate the total electromagnetic fields reflecting and diffracting off the surface of a 2D wedge. Assuming an infinitesimally thin plane wave pulse (or "impulse") excitation incident on the wedge, it was attempted to use the observed impulse response at a fixed position to characterize the wedge as a linear time invariant system (e.g. represent it by a computationally simpler transfer function or equivalent circuit).

The Time Domain Vector Fitting (TDVF) algorithm has been applied in this section to generate the transfer function which leads to the equivalent circuit or "macro model" for the time domain diffraction coefficient. In short, TDVF implements a least squares solution of the impulse response coefficients that best match the observation with a set of exponential terms for its basis (where each exponential term evaluated for all discrete time steps is a vector in the basis).

The objective of this section is then to implement TDVF to obtain an analytical

circuit expression for the time history of the impulse response of a 2D infinite wedge. To begin, it is necessary to prove that Makhoul's equations may be used to calculate the impulse response of a wedge. In their original form, they were derived only to calculate the heaviside step response of a wedge.

5.2 Interactive Visualization of Makhoul's Solution

According to Makhoul's derivation [29], the time history of the $\overrightarrow{E_z}$ or $\overrightarrow{H_z}$ field may be calculated using Eq. (5.1a) through Eq. (5.1c).

$$g(\theta') = \frac{1}{\pi} \tan^{-1} \left[\frac{\sinh(\kappa\tau)\sin(\kappa\pi)}{\cosh(\kappa\tau)\cos(\kappa\pi) - \cos(\kappa\theta')} \right] \tag{5.1a}$$

$$\overrightarrow{H_z} = g(\theta - \theta_0) + g(\theta + \theta_0) \tag{5.1b}$$

$$\overrightarrow{E_z} = g(\theta - \theta_0) - g(\theta + \theta_0) \tag{5.1c}$$

Where the fields at an incident angle, θ_0, upon the wedge of radian angle, α, and for $\kappa = \frac{\pi}{2\pi - \alpha}$, are observed at a radius and radian angle in cylindrical coordinates, (r, θ). The time dependency is parameterized by the variable $\tau = \cosh^{-1}(c \cdot t/r)$, where c is the speed of light and t is the actual time since impact of the plane wave with the wedge. It is obvious that τ is unitless and that all calculations are relative to the observer's radius distance, r. It is therefore convenient to let $c = 1$ for numerical simplicity and when absolute measurements of time, t, are necessary, the true time is given by dividing by the true value for the speed of light, $t_{actual} = c \cdot t/299792458$.

The differential form of Makhoul's equations are then derived analytically by differentiating Eq. (5.1b) and Eq. (5.1c) with respect to time. The resulting formula for the differential fields is given in Eq. (5.2a) through Eq. (5.2c).

$$f_a(\theta') = \frac{(k\sin(\pi k))\cosh\left(k\cosh^{-1}\left(\frac{t}{r}\right)\right)}{\left(\cos(k(\theta')) - \cos(\pi k)\cosh\left(k\cosh^{-1}\left(\frac{t}{r}\right)\right)\right)}$$

$$f_b(\theta') = \frac{\sinh^2\left(k\cosh^{-1}\left(\frac{t}{r}\right)\right)}{\left(\cos(k(\theta')) - \cos(\pi k)\cosh\left(k\cosh^{-1}\left(\frac{t}{r}\right)\right)\right)^2}$$

$$f(\theta') = \frac{fa(\theta') + fb(\theta')(\sin(\pi k)(k\cos(\pi k)))}{\pi\left(fb(\theta')\sin^2(\pi k) + 1\right)} \frac{1}{r \cdot \sqrt{\frac{t^2}{r^2} - 1}} \tag{5.2a}$$

$$\overrightarrow{H_z} = f(\theta - \theta_0) + f(\theta + \theta_0) \tag{5.2b}$$

$$\overrightarrow{E_z} = f(\theta - \theta_0) - f(\theta + \theta_0) \tag{5.2c}$$

An interactive graphical user interface (GUI) was developed to allow rapid tests to be performed on the time history of the wedge impulse. Upon initially loading, the original step response is calculated using Eq. (5.1c) is shown, as in Figure 5.1. The $\overrightarrow{H_z}$ field may be calculated simply by right-clicking within the simulation space and selecting the option for calculating Eq. (5.1b). A fully white color signifies the maximum amplitude of usually 1 (it is 2 when $\overrightarrow{H_z}$ is plotted since the reflected and incident fields are coherent outside the shadow region) whereas gray (the color beneath the beige wedge) signifies zero and black signifies an amplitude of -1.

The TDVF result is shown on the right-hand-side of the window which is also customizable by the user (the scales may be edited by right clicking on the graphs). The vector-fit and time history is shown in the upper right corner of the GUI where-as the point-by-point percent error is shown in the bottom right corner of the GUI.

Among several other options, r, θ, and θ_0 may all be changed simply by left clicking and dragging the red locater spots (to change r) or right-clicking and dragging them

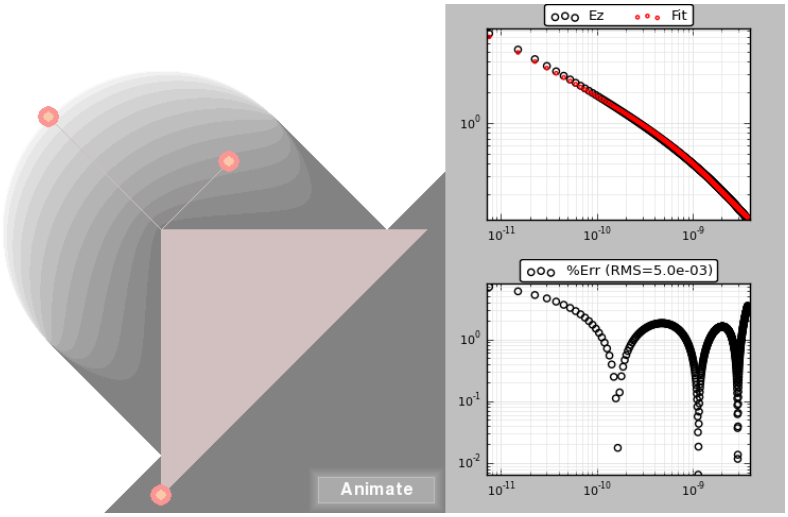

Figure 5.1: User interface developed for interactively plotting the time history of a *diffraction bubble* observed in the interaction of a TE impulse plane wave with an infinite 2D wedge.

(to change an angle value). This allows any case of observation for field incidence and wedge configuration as shown in Figure 5.2.

5.3 Verification of the Differential Form of Makhoul's Eqn.

The differential form of Makhoul's equations of Eq. (5.2a) through Eq. (5.2c) may be verified as correct with a trivial experiment. Since it is well-known that the continuous time impulse function is the derivative of the continuous time heaviside step function, this derivative can be approximately reproduced with a discrete time central differences implementation. This can be done using two heaviside unit step functions that are of equal and opposite amplitude and offset in time by a finite time difference, dt.

By the definition of the continuous time derivative, as $dt \longrightarrow 0$ the result should converge with the analytical expressions of Eq. (5.2a) through Eq. (5.2c). This is numerically verified by shrinking the time offset between the two heaviside unit step functions and dividing their sum by dt. Since the step functions have opposite sign,

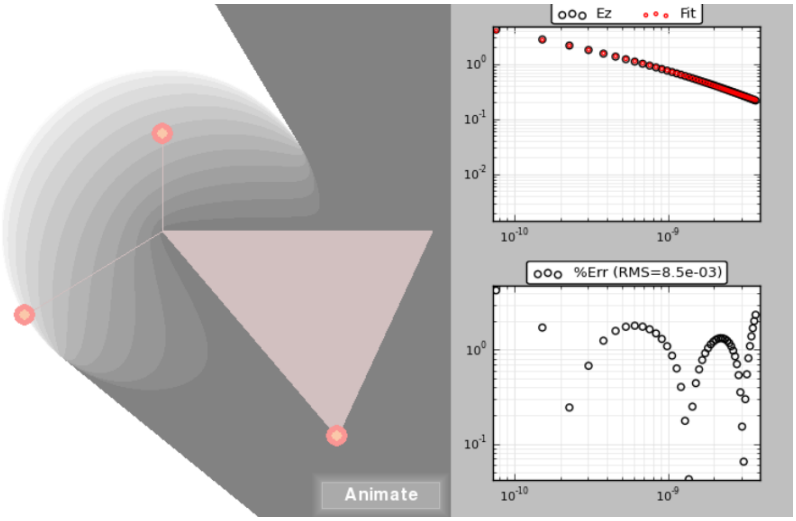

Figure 5.2: User interface developed for interactively plotting the time history of a *diffraction bubble* observed in the interaction of a TE impulse plane wave with an infinite 2D wedge.

the sum is really a subtraction.

Figure 5.3 shows the result of two heaviside step functions spaced by $c \cdot t = 0.01$. The step functions are effectively *chasing* one another in time and are so close that all that is visible to the observer are lines which have thickness equal to the offset in time between the two. Because $\overrightarrow{E_z}$ is plotted, the reflected plane wave impulses have amplitude -1 (black) and incident have amplitdue of +1 (white).

Figure 5.4 shows the actual differential $\overrightarrow{E_z}$ field according to Eq. (5.2c), which matches the central differences result precisely and concludes that the derived expression is correct.

The same procedure is applied for verifying that the $\overrightarrow{H_z}$ fields are correctly calculated for an impulse excitation as shown in Figure 5.5 and Figure 5.6.

The impulse response of an infinite 2D wedge has been shown to be correctly calculated by verifying the observed time history matches that of an equivalent numerical differentiation of the heaviside unit step excitation results. This concludes that the result obtained by TDVF can provide an analytical solution to the impulse

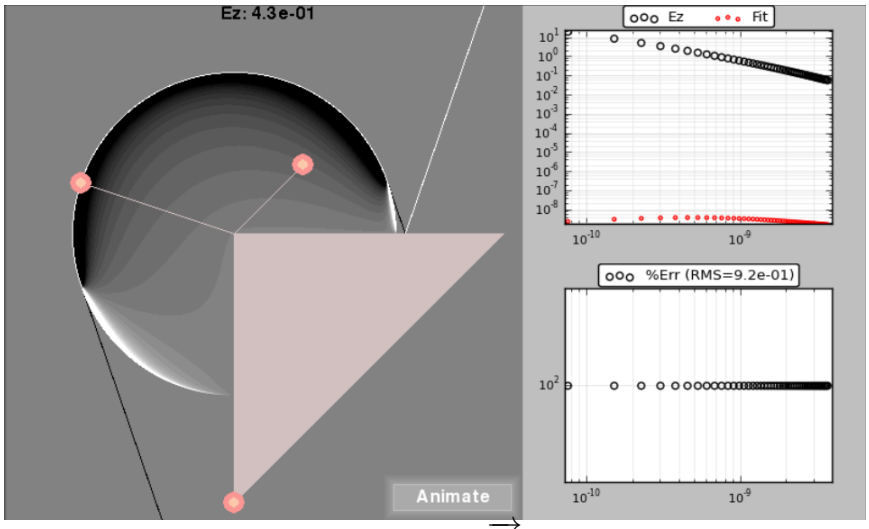

Figure 5.3: Central differences result of the $\overrightarrow{E_z}$ fields calculated by the summation of two equal and opposite heaviside step functions space apart by $c \cdot t = 0.01$.

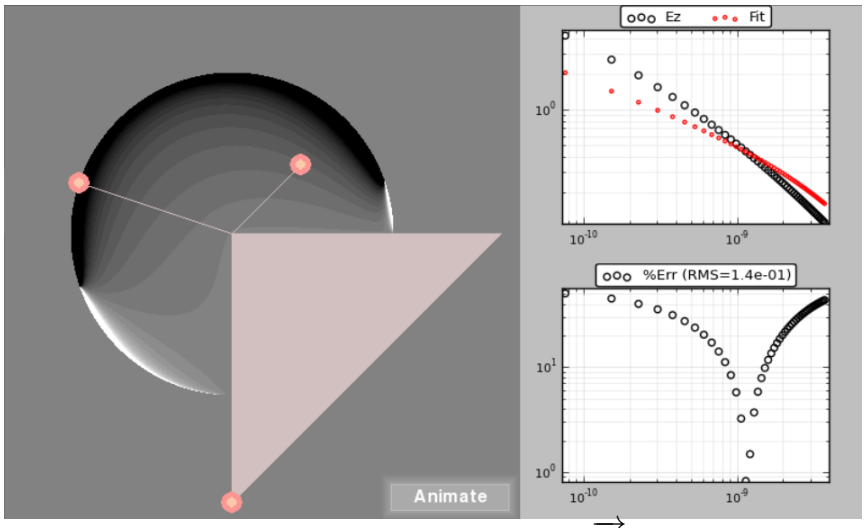

Figure 5.4: Analytical result for the time history of $\overrightarrow{E_z}$ fields caused by an impulse excitation.

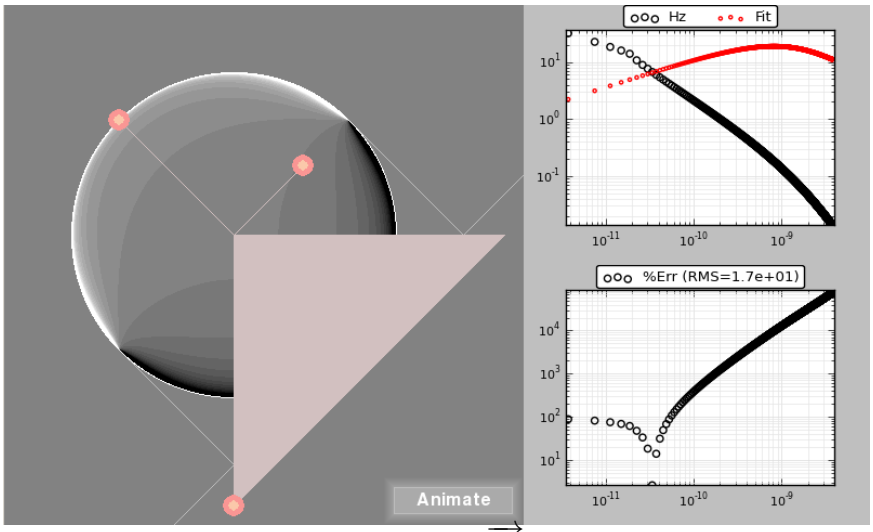

Figure 5.5: Central differences result of the $\overrightarrow{H_z}$ fields calculated by the summation of two equal and opposite heaviside step functions space apart by $c \cdot t = 0.01$.

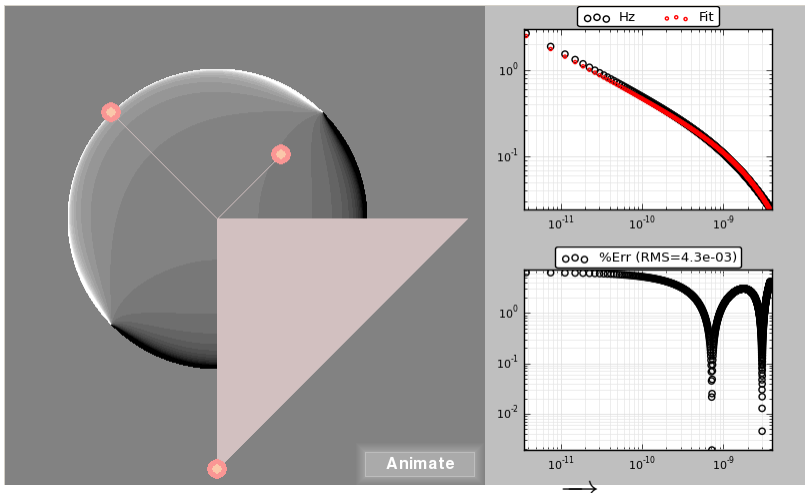

Figure 5.6: Analytical result for the time history of $\overrightarrow{H_z}$ fields caused by an impulse excitation.

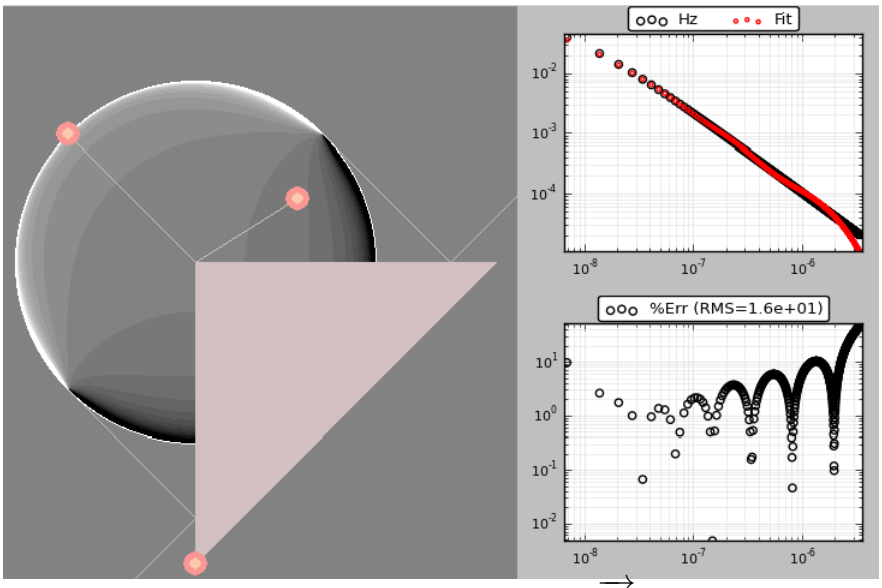

Figure 5.7: TDVF result for large time history of $\overrightarrow{H_z}$ fields. The fit is shown to deviate and have periodic error fluctuation that is not within machine precision as expected.

response of the 2D finite wedge that is far less complicated to implement than the derived forms of Eq. (5.2a) through Eq. (5.2c).

However, the usual basis for the TDVF method does not appear to span the time history measurement. This is shown by Figure 5.7 for the time history of a TE impulse excitation on the wedge. As the time history domain is made larger (observing more time passing by), the TDVF result begins to deviate at late time measurements. Furthermore, the error in the fit is quite high (as indicated in the bottom right corner of Figure 5.7). The error is expected to be within machine precision since there is no noise in the simulation. However not only is it substantially higher, but it has a periodic tendency.

This periodic, fluctuating error implies an over-fit and that the assumed basis may not actually span the vector space implied by Makhoul's equations. To remedy this problem, in the next section a more appropriate basis for Makhoul's equations is found.

5.4 Wedge Time History Basis

To find the right basis to use when fitting the time history of the wedge impulse response, the time dependent terms are collected in order to reduce Makhoul's equations to their simplest possible form. First, notice there is a singularity in Eq. (5.2a) consisting of Eq. (5.3).

$$r \cdot \sqrt{\frac{t^2}{r^2} - 1} \tag{5.3}$$

This causes an infinity in the data which must be removed due to the finite precision in computing the time history. However, it is a relatively small term that is trivial to scale the result by after the fact (so that the singularity is factored back into the final result).

Eq. (5.2a) can be further simplified by applying the definition of the hyperbolic inverse cosine.

$$\cosh^{-1}\left(\frac{t}{r}\right) = \ln\left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right), \quad \frac{t}{r} \geq 1 \tag{5.4}$$

Therefore, applying Eq. (5.4) allows the hyperbolic terms to reduce to algebraic expressions. Applying euler's identity for hyperbolic cosine, all factors of t are expressed in algebraic terms rather than as arguments to trigonometric functions.

$$\cosh\left(k \cdot \cosh^{-1}\left(\frac{t}{r}\right)\right) =$$

$$= \frac{\exp\left\{\kappa \cdot \ln\left[\left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)\right]\right\} + \exp\left\{-\kappa \cdot \ln\left[\left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)\right]\right\}}{2} \tag{5.5}$$

$$= \frac{\left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)^{\kappa} + \left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)^{-\kappa}}{2}$$

Since there is only interest in obtaining time histories (functions of t), all other parameters can be assumed as constants if they can be collected without factors of t involved. With this simplification, the only variables remaining in Eq. (5.2a) can be reduced to Eq. (5.6a) and Eq. (5.6b):

$$A = \left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)^{\kappa} + \left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)^{-\kappa} \tag{5.6a}$$

$$A' = \left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)^{\kappa} - \left(\frac{t}{r} + \sqrt{\left(\frac{t}{r}\right)^2 - 1}\right)^{-\kappa} \tag{5.6b}$$

Where, by Eq. (5.7), Eq. (5.2a) becomes a function of only one variable, A.

$$\begin{aligned} A^2 &= \left(\left(\frac{t}{r} + \sqrt{\frac{t}{r} - 1} \times \sqrt{\frac{t}{r} + 1}\right)^{\kappa} + \left(\frac{t}{r} + \sqrt{\frac{t}{r} - 1} \times \sqrt{\frac{t}{r} + 1}\right)^{-\kappa}\right)^2 \\ &= \frac{\left(\left(\frac{t}{r} + \sqrt{\frac{t}{r} - 1} \times \sqrt{\frac{t}{r} + 1}\right)^{2\kappa} + 1\right)^2}{\left(\frac{t}{r} + \sqrt{\frac{t}{r} - 1} \times \sqrt{\frac{t}{r} + 1}\right)^{2\kappa}} \\ &= \frac{1}{\left(\frac{t}{r} + \sqrt{\frac{t}{r} - 1} \times \sqrt{\frac{t}{r} + 1}\right)^{2\kappa}} + \left(\frac{t}{r} + \sqrt{\frac{t}{r} - 1} \times \sqrt{\frac{t}{r} + 1}\right)^{2\kappa} + 2 \\ &= (A')^2 + 4 \end{aligned} \tag{5.7}$$

Therefore $A' = \sqrt{A^2 - 4}$ and is thus dissolved. The additional trigonometric functions that do not carry factors of t in their arguments have already been declared as constants in the basis since the objective is to parametrize the result as a function

of only time at first. Therefore, after reducing them to their respective euler identities, is further simplified to .

$$h_a(A,C) = \frac{A}{4\left[1/2\left(C+\frac{1}{C}\right) - 1/4\left(B+\frac{1}{B}\right)A\right]}$$

$$h_b(A,C) = \frac{\left(B+\frac{1}{B}\right)\left(A^2-4\right)}{16\left[1/2\left(C+\frac{1}{C}\right) - 1/4\left(B+\frac{1}{B}\right)A\right]^2}$$

$$h(A,C) = \frac{j\kappa\left(B-\frac{1}{B}\right)\left(h_a(A,C)+h_b(A,C)\right)}{\pi\left(1-h_b(A,C)\right)}\frac{1}{r\cdot\sqrt{\frac{t^2}{r^2}-1}} \tag{5.8a}$$

$$\overrightarrow{H_z} = h(A, C=e^{-j\kappa(\theta-\theta_0)}) + h(A, C=e^{-j\kappa(\theta+\theta_0)}) \tag{5.8b}$$

$$\overrightarrow{E_z} = h(A, C=e^{-j\kappa(\theta-\theta_0)}) - h(A, C=e^{-j\kappa(\theta+\theta_0)}) \tag{5.8c}$$

Where $B = e^{-j\kappa\pi}$. The constants, B and C, can then be lumped together as numerator, n_p, and denomenator, d_q, coefficients of a rational function form of Eq. (5.8a). Thus, both the electric and magnetic field amplitudes of the wedge impulse response can be spanned by the rational polynomial basis function of Eq. (5.9).

$$\overrightarrow{E_z}, \overrightarrow{H_z} \Rrightarrow \frac{n_0 + An_1 + A^2n_2 + A^3n_3}{d_0 + Ad_1 + A^2d_2 + A^3d_3 + A^4d_4}\frac{1}{r\cdot\sqrt{\frac{t^2}{r^2}-1}} \tag{5.9}$$

This rational polynomial form can be expanded into partial fractions, allowing the poles in this *nonlinear* time scale, A, to be viewed explicitly. Once the poles and their residues are realized, an equivalent circuit may be generated.

What this basis implies however is not entirely obvious at first, since the time domain expression is more similar to a frequency domain transfer function (it is a

rational polynomial). Normally a time domain impulse response is typically characterized by a sum of decaying exponential terms. Therefore, this logic implies the frequency response of a circuit given by the modal parameters of Eq. (5.9) will yield the actual time domain impulse response by directly assuming discrete values of A as frequency samples rather than time samples.

5.5 Numerical Implementation of Makhoul Basis

Eq. (5.9) is used as the basis of the original Frequency Domain Vector Fitting algorithm (FDVF). FDVF will then identify the poles and residues of the partial fraction expansion of Eq. (5.9), assuming discrete values of A are frequency samples. The FDVF result is verified by analytically calculating the partial fraction expansion of Eq. (5.9). This cannot be done symbolically as it requires the numerical computation of the roots of the function for given values of the wedge angle, incident angle and observer position.

Consider the time history, $1 < t < 100$, of the impulse response for a TE wave of incidence, $\theta_0 = 45°$, upon a wedge angle of $\alpha = 90°$, at an observer position of $r = 0.9$, and $\theta = 45°$. The factorisable rational polynomial form of $\overrightarrow{H_z}$ is then given by evaluating the coefficients of Eq. (5.9) for the given parameters above to yield .

$$\overrightarrow{H_z} = \frac{A^2\left(-7.3 + j12.6\right) + \left(29.0 - j50.3\right)}{A^4\left(-19.7 + j34.2\right) + A^2\left(98.7 - j170.9\right) + \left(-79.0 + 136.8\right)} \frac{1}{r \cdot \sqrt{\frac{t^2}{r^2} - 1}}$$

Where the coefficients have been rounded for simplicity. The numerator and denominator can then be factored and a partial fraction expansion is found to be:

$$\overrightarrow{H_z} = \left(\frac{0.184}{A-1} - \frac{0.184}{A+1}\right) \frac{1}{r \cdot \sqrt{\frac{t^2}{r^2} - 1}} \tag{5.10}$$

Where now the residues of Eq. (5.10) have been rounded for simplicity. As a

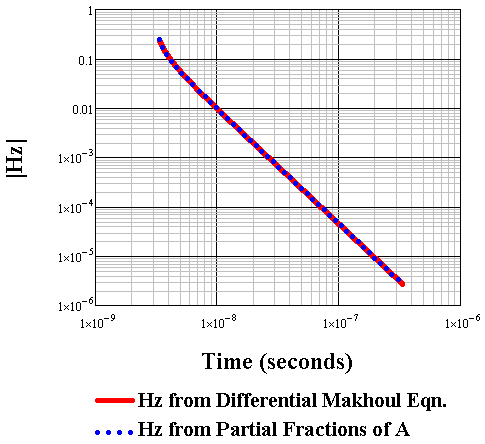

Figure 5.8: Analytical expression for H_z using both the original expression for the impulse response of Eq. (5.2a) as well as the partial fraction expansion of Eq. (5.9).

simple verification, Eq. (5.10) and Eq. (5.2a) are compared in Figure 5.8, illustrating that the partial fraction expansion suffered no significant round-off error when done numerically computed; it is an exact solution.

This rational polynomial basis (which is intuitively very similar to the frequency domain version of Vector Fitting) is then implemented and the exact poles and resides of Eq. (5.10) are found precisely. This was also repeated for several variations of incident angle, wedge angle, and observer position. A few of the variations are shown in Figure 5.9 through Figure 5.11.

Judging by the precise accuracy of the fit it seems Makhoul's equations may not be expanded into a finite sum of exponential terms. Instead, it appears a power series expansion would apply. This can be intuitively verified by fitting the curve much like in a Bode procedure of curve fitting. On a log-log plot with the diffraction bubble singularity of Eq. (5.3) removed, the time history forms a straight line. The slope of this line approximately corresponds to a power function exponent. The slope of the time history for the case of Figure 5.9 is shown in Figure 5.12.

A single power function is a fairly good approximation of the time history, and suggests that more phenomenon must be in the observation other than what can be

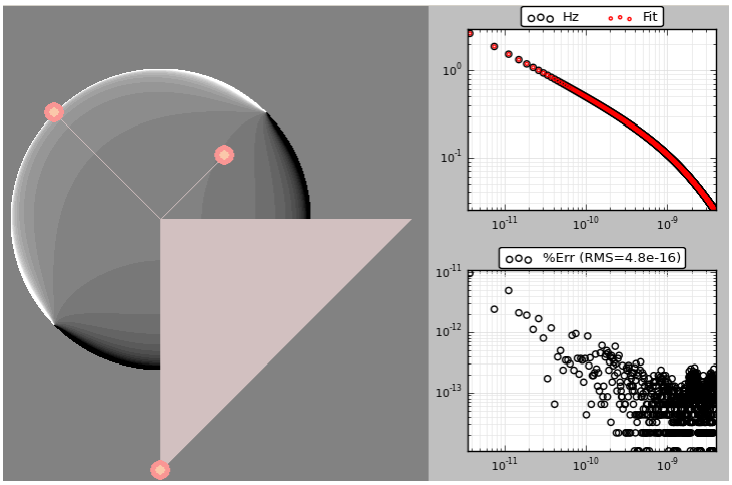

Figure 5.9: FDVF result using Makhoul basis of Eq. (5.10) for H_z time history of 90° wedge, incident angle of 135°, and observer position of $r = 0.9$, $\theta = 45°$.

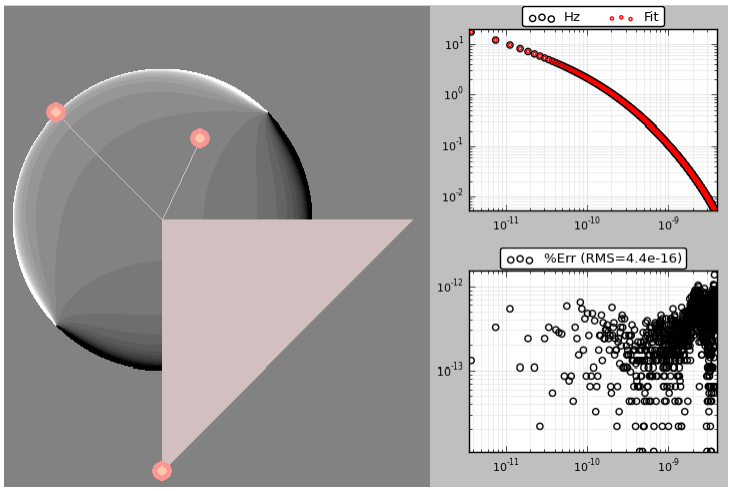

Figure 5.10: FDVF result using Makhoul basis of Eq. (5.10) for H_z time history of 90° wedge, incident angle of 135°, and observer position of $r = 0.9$, $\theta = 65°$.

explained by decaying exponential terms.

The reason why the time history may be explained in terms of power functions has been conjectured before [30]. The notion is that charges present on the wedge surface are unbounded by the infinite wedge cross section. They move from the tip of the wedge where diffraction begins onwards to infinity and thus the time history is that of a constantly evolving charge distribution. However, passive circuit components are finite and are time invariant which is why a finite number of circuit elements cannot

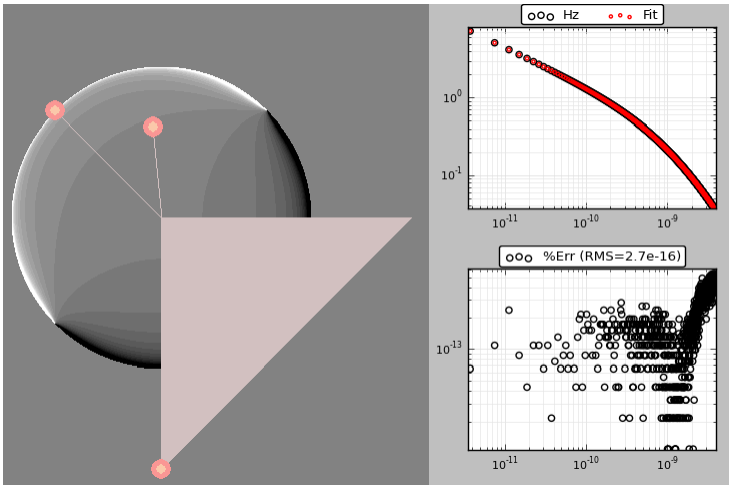

Figure 5.11: FDVF result using Makhoul basis of Eq. (5.10) for H_z time history of 90° wedge, incident angle of 135°, and observer position of $r = 0.9$, $\theta = 95°$.

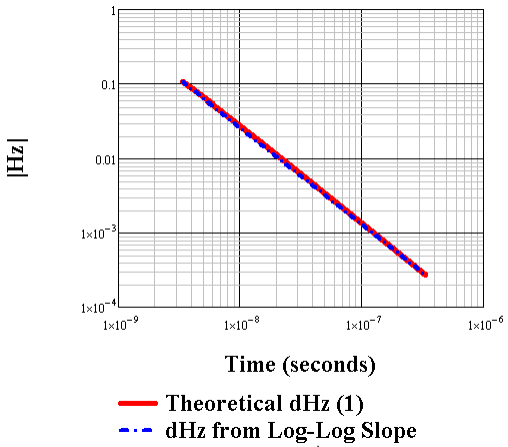

Figure 5.12: Plotting the exact solution for $\overrightarrow{H_z}$ on a log-log plot for $r = 0.9$, $\alpha = 90°$, $\theta = 45°$, $\theta_0 = 135°$. The approximate power function is found from the slope to be $\overrightarrow{H_z} \approx 1.05 \cdot 10^{-12} \left(\frac{t}{c}\right)^{-1.3}$

possibly reproduce this phenonmenon.

Though an equivalent circuit can not be formed from the derived Makhoul basis, it does not argue against there being a circuit equivalent to arbitrary perfect electrically conducting scatterers. Rather, this points out one limitation of the goal in mind that infinitely sized scatterers may require additional consideration to acheive a circuit equivalent. Furthermore, it may be more pragmatic to only work with finite-sized scatterers.

REFERENCES

[1] A. Semlyen and B. Gustavsen, "Rational approximation of frequency domain responses by vector fitting," *IEEE Transactions on Power Delivery*, vol. 14, no. 3, pp. 1052–1061, 1999.

[2] A. S. N. H. Oppenheim, Alan V. Willsky, *Signals and Systems*, S. Hamid, Ed. Pearson Education, 1998.

[3] A. Nicolson and G. F. Ross, "Measurement of the intrinsic properties of materials by time-domain techniques," *Instrumentation and Measurement, IEEE Transactions on*, vol. 19, no. 4, pp. 377–382, 1970.

[4] W. B. Weir, "Automatic measurement of complex dielectric constant and permeability at microwave frequencies," *Proceedings of the IEEE*, vol. 62, no. 1, pp. 33–36, 1974.

[5] L. De Tommasi, D. Deschrijver, and T. Dhaene, "Single-Input-Single-Output Passive Macromodeling via Positive Fractions Vector Fitting," in *Signal Propagation on Interconnects, 2008. SPI 2008. 12th IEEE Workshop on*, 2008, pp. 1–2.

[6] B. Gustavsen and A. Semlyen, "Enforcing passivity for admittance matrices approximated by rational functions," *Power Systems, IEEE Transactions on*, vol. 16, no. 1, pp. 97–104, 2001.

[7] A. Semlyen and B. Gustavsen, "Vector fitting by pole relocation for the state equation approximation of nonrational transfer matrices," *Circuits, Systems and Signal Processing*, vol. 19, no. 6, pp. 549–566, 2000.

[8] D. Deschrijver, B. Haegeman, and T. Dhaene, "Orthonormal Vector Fitting: A Robust Macromodeling Tool for Rational Approximation of Frequency Domain Responses," *Advanced Packaging, IEEE Transactions on*, vol. 30, no. 2, pp. 216–225, 2007.

[9] B. Gustavsen, "Relaxed vector fitting algorithm for rational approximation of frequency domain responses," in *Signal Propagation on Interconnects, 2006. IEEE Workshop on*, 2006, pp. 97–100.

[10] E. F. Knott, J. F. Shaeffer, and M. T. Tuley, *Radar Cross Section (2nd Edition)*. SciTech Publishing, 2004.

[11] R. E. Diaz and N. G. Alexopoulos, "The Application of the Analytic Continuation Model of Dispersive Materials to Electromagnetic Engineering*," *Electromagnetics*, vol. 18, no. 4, pp. 395–422, Jul. 1998.

[12] R. Diaz, J. Peebles, R. Lebaron, Z. Zhang, and L. Lozano-Plata, "United states patent 7889148 b2," 2008.

[13] D. Pozar, *Microwave Engineering*. Wiley, 1989.

[14] S. Hwang, S. Min, M. Swaminathan, V. Venkatakrishnan, H. Chan, F. Liu, V. Sundaram, S. Kennedy, D. Baars, B. Lacroix, Y. Li, and J. Papapolymerou, "Characterization of Next Generation Thin Low-K and Low-Loss Organic Dielectrics From 1 to 110 GHz," *Advanced Packaging, IEEE Transactions on*, vol. 33, no. 1, pp. 180–188, 2010.

[15] M. Swaminathan and Ege Engin, *Power Integrity Modeling and Design for Semiconductors and Systems*. Prentice Hall, 2007.

[16] S. A. Schelkunoff, "Some Equivalence Theorems of Electromagnetics and Their Application to Radiation Problems," *Bell System Technical Journal*, vol. 15, no. 1, pp. 92–112, 1936.

[17] Y. Guo, *M.S. Thesis: Design of a Compact High Fidelity Free-Space Material Measurement Test Set-Up*. Arizona State University, 2005.

[18] R. A. Waldron, "Perturbation theory of resonant cavities," *Proceedings of the IEE - Part C: Monographs*, vol. 107, no. 12, pp. 272–274, 1960.

[19] ——, "Theory of a Strip-Line Cavity for Measurement of Dielectric Constants and Gyromagnetic-Resonance Line-Widths," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 12, no. 1, pp. 123–131, 1964.

[20] R. A. Waldron and S. P. Maxwell, "Note on the Measurement of Material Properties by the Strip-Line Cavity (Correpsondence)," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 13, no. 5, p. 711, 1965.

[21] R. E. Diaz, W. M. Merrill, and N. G. Alexopoulos, "Analytic framework for the modeling of effective media," *Journal of Applied Physics*, vol. 84, no. 12, 1998.

[22] S. P. Maxwell, "Strip-line cavity resonator for measurement of magnetic and dielectric properties of ferrites at low microwave frequencies," *Marconi Review*, vol. 27, pp. 22–29, 1964.

[23] E. Pardo, D.-X. Chen, and A. Sanchez, "Demagnetizing factors for square bars," *Magnetics, IEEE Transactions on*, vol. 40, no. 3, pp. 1491–1498, 2004.

[24] D.-X. Chen, J. A. Brug, and R. B. Goldfarb, "Demagnetizing factors for cylinders," *Magnetics, IEEE Transactions on*, vol. 27, no. 4, pp. 3601–3619, 1991.

[25] A. Aharoni, "Demagnetizing factors for rectangular ferromagnetic prisms," *Journal of Applied Physics*, vol. 83, no. 6, pp. 3432–3434, 1998.

[26] C. A. Jones, "Permittivity and permeability measurements using stripline resonator cavities-a comparison," *Instrumentation and Measurement, IEEE Transactions on*, vol. 48, no. 4, pp. 843–848, 1999.

[27] T. Sebastian, S. A. Clavijo, and R. E. Diaz, "Improved accuracy thin film permeability extraction for a microstrip permeameter," *Journal of Applied Physics*, vol. 113, no. 3, pp. –, 2013.

[28] D. Deschrijver, T. Dhaene, and D. De Zutter, "Robust parametric macromodeling using multivariate orthonormal vector fitting," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 56, no. 7, pp. 1661–1667, 2008.

[29] J. I. Makhoul, "Contour plots for the diffraction of a pulse by a wedge," *Radio Science*, vol. 1, no. 5, pp. 609–613, 1966.

[30] J. H. Thompson, "Closed Solutions for Wedged Diffraction," *SIAM Journal on Applied Mathematics*, vol. 22, no. 2, pp. pp. 300–306, 1972.

[31] F. Liu and T. T. Y. Wong, "Bistatic impulse responses of a perfectly conducting sphere," in *Antennas and Propagation Society International Symposium, 1994. AP-S. Digest*, vol. 3, 1994, pp. 2329–2332 vol.3.

[32] C. L. Bennett, W. L. R. Weeks, and P. U. S. of Electrical Engineering, *A Technique for Computing Approximate Electromagnetic Impulse Response of Conducting Bodies*, ser. TR-EE. School of Electrical Engineering, Purdue University, 1968.

[33] S. H. Talisa, "Application of Davidenko's Method to the Solution of Dispersion Relations in Lossy Waveguiding Systems," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 33, no. 10, pp. 967–971, 1985.

[34] B. Carnahan and H. Luther, *Applied Numerical Methods*. Wiley, 1969.

[35] A. Ralston, "Runge-kutta methods with minimum error bounds," *Mathematics of Computation*, vol. 16, p. 437, 1962.

[36] S. C. Chapra and R. Canale, *Numerical Methods for Engineers*, 5, Ed. New York: McGraw-Hill, 2006.

APPENDIX A

DAVIDENKO'S METHOD FOR PERMITTIVITY EXTRACTION

In this work, it was necessary to program a method for solving the permittivity of material under test in an admittance tunnel following the procedure in [17, 33]. A numerical method was written using OpenCL (to run on a graphics processing unit) since no similar solver libraries were known to exist yet for that language. Davidenko's Method (DM) was used in place of the well known Newton-Raphson Methods (NRM) or *Newton's method for nonlinear systems*, since these methods can suffer from stability issues. DM is very similar to NRM, except it calculates the increment towards the root in differential form. This adds complexity since an ordinary differential equation (ODE) solver is then necessary to form the best approximation of that slope over a discrete interval step. A Runge-Kutta-Fehlberg method (RKF) was chosen to perform this task given its high accuracy and acceptable stability. Dormand-Prince coefficients (RK5(4)7M) were used which have been shown to have the smallest global error relative to other RKF methods [34].

This section begins by introducing DM through comparisons with NRM, which it is derived from. It starts with the simple example of solving a root for one equation and extends to a system of equations. Stability issues with NRM are outlined and justify the use of DM for problems with shallow slopes symmetry about the x-axis. In summary, DM is a differential form of NRM.

A.1 Davidenko's Method

The NRM is a slope-following root finder. The algorithm shifts an approximation (initially guessed by the user), r_k, to a closer approximation, r_{k+1}, of the function's root. Each iteration corrects r_k by subtracting the run of the function's slope at r_k. For the one-dimensional case, the closer approximation is found by:

$$r_{k+1} = r_k - \frac{f(r_k)}{f'(r_k)}$$

Geometrically, r_{k+1} is the zero-crossing of the line tangent to the function's slope at r_n. When r_{k+1} is sufficiently close to the zero-crossing of the actual function, the method is finished.

Finding one-dimensional roots numerically can be solved with derivative-free algorithms such as secant and bisection methods (especially for linear functions). But for N dimensions and an algebraic set of $M = N$ non-linear functions, slope-following calculations such as NRM become necessary means of Finding roots since multivariate functions become too irregular to solve with computationally cheaper methods.

Let x_n represent a variable of the nth dimension out of the N total dimensions (with iteration k described below). Also, let $f_m(x_1,\ x_2,\ x_3,\ x_n,\ \ldots.\ x_N$ represent the mth function out of the set of $M = N$ functions. These are written in matrix form as:

$$\mathrm{X}_{n,k} = \begin{bmatrix} x_{n,k} & \cdots & x_{n,k+1} \\ \vdots & \ddots & \vdots \\ x_{N,k} & \cdots & x_{N,k+1} \end{bmatrix}$$

$$\mathrm{F}_m = \begin{bmatrix} f_m(x_n .. \cdot \ldots \cdot \cdots x_N) = 0 \\ \vdots \\ f_M(x_n .. \cdot \ldots \cdot \cdots x_N) = 0 \end{bmatrix}$$

In progressing from one dimension to higher dimensions, the partial derivative of each function, $f_m(x_n, \ \ldots, \ x_N)$, is calculated with respect to each x_n. Each dimension recursively improves the initial guess of every dimension, n. After each iteration of recursion, k, the root of each dimension is corrected towards the next closest approximation – just like in the one-dimensional case. This operation conducted for all M functions is:

$$\mathrm{F}_m(\mathrm{X}_{k+1}) = \mathrm{F}_m(\mathrm{X}_k) + \sum_{n=1}^{N} \frac{\partial \mathrm{F}_m}{\partial \mathrm{X}_{n,k}} \ (\mathrm{X}_{n,k+1} - \mathrm{X}_{n,k}) = 0 \tag{A.1}$$

The partial derivative operation for all iterations of m, n can be formulated in matrix notation as the Jacobian matrix, J:

$$J_{m,n} = \begin{bmatrix} \dfrac{\partial F_M}{\partial x_n} & \cdots & \dfrac{\partial F_M}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial F_M}{\partial x_n} & \cdots & \dfrac{\partial F_M}{\partial x_n} \end{bmatrix}$$

Eq. (A.1) can then be rewritten as:

$$\mathrm{F}(\mathrm{X}_{k+1}) = \mathrm{F}(\mathrm{X}_k .) + \mathrm{J} \cdot (\mathrm{X}_{k+1} - \mathrm{X}_k) = 0$$
$$\mathrm{F}(\mathrm{X}_k) + \mathrm{J} \cdot (\mathrm{X}_{k+1} - \mathrm{X}_k) = 0$$

then,

$$\mathrm{X}_{k+1} = -\mathrm{J}^{-1} \cdot \mathrm{F} + \mathrm{X}_\mathrm{k} \tag{A.2}$$

Eq. (A.2) can also be rewritten to avoid sub-scripting the algorithm's kth and $kth+1$ iteration by defining:

$$\Delta \mathrm{X} = \mathrm{X}_{k+1} - \mathrm{X}_k \tag{A.3}$$

then, a simple substitution of equation Eq. (A.2) into Eq. (A.3) reduces it to:

$$\Delta \mathrm{X} = -\mathrm{J}^{-1} \cdot \mathrm{F} \tag{A.4}$$

Equation Eq. (A.4) is the step of NRM for solving the root of a set of equations. There are stability issues related to this method which can motivate the use of other numerical methods. For instance, when an initial guess is made near a horizontal asymptote, the division by the vanishing slope can hyper-correct ($\Delta \mathrm{X} \gg \mathrm{X}_k$) , if

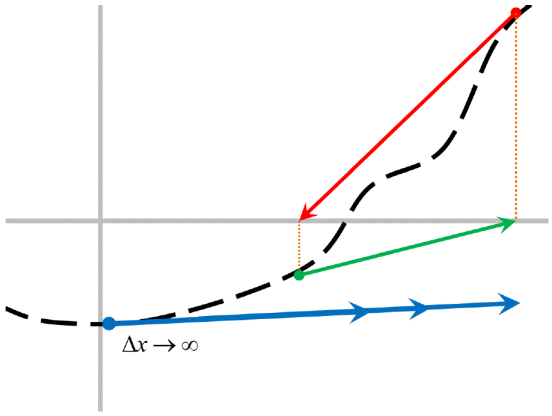

Figure A.1: The blue arrow signifies when the slope is nearly zero, hyper-correcting the guess, $\triangle x$. The red and green arrows illustrate non-convergent cycles where both slopes direct t1$_1$e correction in opposing directions, resulting in an infinite loop.

not destabilize equations Eq. (A.1) and Eq. (A.4). Also, the algorithm may non-convergently cycle about regions with slopes that correct the next guess back to where it previously was. Figure 1 illustrates these two situations.

Therefore, the success of this method relies on the initial guess locating itself away from these caveats. Since that can be impractical when guessing is hard, one way to further ensure ΔX stays small is to scale F to be small, as in Figure *2*. Scaling F by a factor, t, prevents ΔX from hyper-correcting, especially as $t \longrightarrow 0$. In summary, this means equation Eq. (A.4) is modified as in the following:

$$\Delta \mathrm{X} \to 0 = -(\mathrm{t} \to 0)\mathrm{J}^{-1} \cdot \mathrm{F}$$

$$\frac{\Delta \mathrm{X} \to 0}{\mathrm{t} \to 0} = -\mathrm{J}^{-1} \cdot \mathrm{F}$$

$$\frac{d\Delta \mathrm{X}}{dt} = -\mathrm{J}^{-1} \cdot \mathrm{F}$$

Since infinitesimal X,

$$\frac{d\mathrm{X}}{dt} = -\mathrm{J}^{-1} \cdot \mathrm{F} \tag{A.5}$$

The differential form of NRM in equation Eq. (A.5) is DM. Where the scaling factor, t, has been demoted to just a dummy variable for the purpose of differentiation. Since the step in Davidenko's method is only a slope of the increment towards a solution, it is necessary to have an ODE solver which efficiently grids the solution space and accurately obtains the slope at each point of interest to calculate the value of the original function.

A.2 Runge Kutta Method

Obtaining an accurate approximation of the function's slope over an interval is a major theme in all one step ODE solvers. The entire solution relies on a recursing summation of the product of the slope and step size over each interval. The solution

could have a better approximated slope if many terms of a Taylor expansion of the ODE about the previous step (t_i) were calculated:

$$y(t_{i+1}) = y(t_i) + hf(t_i, y_i) + \frac{h^2}{2!}f'(t_i y_i) + O(h^3) \tag{A.6}$$

But this requires calculating derivatives higher than the ODE, $y'(t_i) = f(t,\ y)$. Instead, we could fit the curve of $y(t)$ by a weighted average of slopes at different points given by the ODE. Then, Eq. (A.6) can be replaced with a different approach:

$$\tilde{y}_{i+1} = \tilde{y}_i + h \cdot \phi(t_i, \tilde{y}_i, h) \tag{A.7}$$

where $\tilde{y}_{i+1}$ is the next approximation and ϕ is referred to as the increment function, which contains the weighted average. The higher order methods contain more derivatives to be averaged. The most intuitive case to gain insight in is the second order methods (a weighted average of two derivatives). Then, Eq. (A.7) looks like:

$$\tilde{y}(t_{i+1}) = \tilde{y}(t_i) + h(p \cdot k_1 + q \cdot k_2)$$

$$= \tilde{y}(t_i) + h \cdot (p \cdot f(t_i,\ y_i) + q \cdot f(t_i + p,\ y_i + q)) \tag{A.8}$$

Where $f(t_i y_i) = y'(t) = k_1$ is a slope given by the ODE and $f(t_i + ph, y_i + qhk_1) = k_2$ is the slope at an intermediate step – quite similar to a multi-step method approach except these coefficients ($p,\ q$) are predetermined by the method rather than based on previous calculations. The coefficients are determined by comparing like terms of $\tilde{y}(t_{i+1})$ and a Taylor polynomial of $y(t)$. First, expand k_2 in two variables and limit the truncation error, as in:

$$f(t_i + ph, y_i + qhk_1) = f(t, y) + ph\frac{\partial}{\partial t}f(t, y) + qh\frac{\partial}{\partial y}f(\mathrm{t}, y) \cdot f(t, y) + O(h^2) \tag{A.9}$$

Then, plugging Eq. (A.9) back into Eq. (A.8) will improve the approximation to $O(h^3)$ truncation error, and the coefficients must be found by equating like terms with a second order Taylor expansion of the function, Eq. (A.6) in two variables:

$$\tilde{y}(t_{i+1}) = \tilde{y}(t_i) + [(\mathrm{a} + \mathrm{b}) \cdot f(\mathrm{t}_i, y_i)]h + [b \cdot p\frac{\partial}{\partial t}f(t, y) + b \cdot q\frac{\partial}{\partial y}f(t, y) \cdot f(t, y)]h^2 + O(h^3)$$

$$\tilde{y}(t_{i+1}) = y(t_i) + 1f(1_i,\ y_i)h - \frac{1}{2}\left[\frac{\partial}{\partial t}f(t, y) + \frac{\partial . f}{\partial y}(t, y) \cdot f(\mathrm{t}, y)\right]h^2 + \mathrm{O}(h^3)$$

Above, the approximation, ($\tilde{y}(t_{i-1})$) based on Eq. (A.8) is compared (note where in red) to a second order Taylor Polynomial. The conclusion is that:

$$\mathrm{a} + \mathrm{b} = 1,\ b \cdot p = \frac{1}{2},\ b \cdot q = \frac{1}{2}$$

Since there are four unknowns and only three equations there are infinitely many solutions. This grants a creative license for assuming one of the coefficients and then

solving for the other three based on that criterion. For instance, using Ralston's Method [35] gives the following form with the assumption:

$$b = \frac{2}{3},\ a = \frac{1}{3},\ p = \frac{3}{4}\ .\ q = \frac{3}{4}$$

$$\tilde{y}_{i+1} = \tilde{y}_i + h(\frac{1}{3}f(\mathrm{f}_i,\ y_i) - \frac{2}{3}f(t_{\dot{2}} + \frac{3}{4}h,\ y_i + \frac{3}{4}hf(\mathrm{f}_i,\ y_i)))$$

Selecting coefficients is done to minimize the local truncation error. For instance, Ralston selected b to match the $O(h^3)$ terms similar to how the above matched the $O(h)$ and $O(h^2)$ terms. This will also constrain the problem, producing a unique b that has been chosen to optimize the truncation error by matching higher order terms [36].