uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

u Ottawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Wei Fang

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)

GRADE / DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Model-Based Retrieval of Phase-Related Quantities for Two-Port Microwave Devices

TITRE DE LA THÈSE / TITLE OF THESIS

D. McNarmara

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

M. Yagoub

Q. J. Zhang

Gary W. Slater

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Model-Based Retrieval of Phase-Related Quantities for Two-Port Microwave Devices

by

Wei Fang

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Master of Applied Science
in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

September 2005

Your file Votre référence
ISBN: 0-494-11268-9
Our file Notre référence
ISBN: 0-494-11268-9

ABSTRACT

In this thesis a new method is proposed for the retrieval of phase-related quantities from magnitude-only measurements on two-port microwave systems. The method is applicable when a microwave system is manufactured in relatively large quantities. A reduced-order model is developed using the prototype of such a system. Such a reduced-order model will include a number of quantities or "coefficients" (other than frequency) whose value is determined using a full vector data set on the prototype. Establishing the form required for the reduce-order model acts as the additional constraints that are needed to provide a unique relationship between the transfer function magnitude and its phase. When later production models are tested using magnitude-only measurements, optimization is used (with the above-mentioned "coefficients" as the optimization variables, for which only small adjustments are needed from one device to another) to match the measured and modeled magnitude responses. The retrieved "coefficients" are then used to predict the phase response using the model. The use of the model-based retrieval technique is demonstrated through its application to three different types of device.

Keywords: Microwave networks, network analyzer measurements, phase retrieval from magnitude data.

ACKNOWLEDGEMENTS

I would like to first thank my supervisor, Dr. Derek A. McNamara. He is the greatest professor and mentor I have known. Secondly I want to thank my family and all my friends, especially Meng Chu, for their support.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 THE CHARACTERISATION OF PASSIVE MICROWAVE CIRCUITS

The situation often occurs where the magnitude and phase response of microwave circuits and sub-systems are required to be verified during production. The various prototypes of such units can be tested using vector network analyzers (VNA), so that both magnitude and phase (and hence secondary quantities derived from the phase response, such as group delay and group delay variation) can be measured directly. However, in production testing, where measurements must consume as little time as possible, and may need to be repeated after various stages of environmental testing, it may be either undesirable or simply difficult to perform VNA testing of all units.

One such situation is that of the testing of an antenna positioning mechanism (APM). Such APMs must be tested in very many different angular positions [2], and the resulting extreme flexing of the VNA measurement cables strongly influences the measured phase response unless great (and time-consuming) care is taken. In these, and many other instances, it is preferable to perform scalar magnitude-only measurements, since the measurement cables that flex then carry only low-frequency "detected" signals whose phase is little affected by physical movement, and the measurements themselves might not require highly-qualified technicians. However, it is then necessary to somehow recover the phase information (i.e. if group delay characteristics are needed) because this

information is indeed a big concern in signal processing (i.e. phase distortion can cause great difficulty in decoding signals in digital modulation). This thesis considers ways of doing this recovery and discusses their effectiveness.

Although the methods to be developed in this thesis can be adapted to multi-port cases we will concentrate on two-port measurements. The passive two-port microwave network in Figure 1.1-1 can be characterized in terms of four scattering parameters $S_{11}(\omega)$, $S_{12}(\omega)$, $S_{21}(\omega)$ and $S_{22}(\omega)$, where $\omega = 2\pi f$ is the angular frequency and f is the frequency. The scattering parameters $S_{ij}(\omega)$ are complex quantities and thus have a magnitude $\left|S_{ij}(\omega)\right|$ and phase $\arg\{S_{ij}(\omega)\}$. A vector network analyzer is able to measure the full set of scattering parameters in both magnitude and phase, outputting these as a set of complex numbers versus frequency. A scalar network analyzer is able to measure the magnitudes only.

1.2 OVERVIEW OF THE THESIS

The rapid development of modeling techniques, both at the equivalent network model and electromagnetic model levels, has greatly benefited the microwave circuit design process [2,3,4,5]. It appears that these have not been exploited at the production line testing level however, in spite of the fact that it is believed by many that "in developing products and systems, testing, and not design, is usually the more expensive, time-consuming, and difficult activity" [6].

An underlying premise of this thesis is the following : When a device or sub-system is in production it will have gone through a design phase where engineers will have used

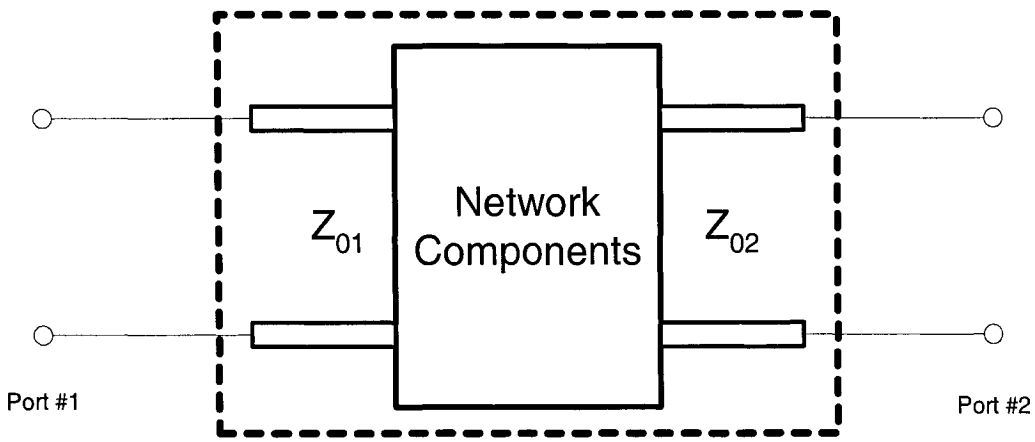


Figure 1.1-1 : Two-Port Microwave Network

various levels of modelling. We maintain that it is advantageous to carry all this valuable information over to the test phase and use it to cut down on the extent of the testing required or to simplify the testing. In particular, we should use all the available information gained during the design phase to allow efficient and reliable retrieval of the phase information from the measured magnitude information.

Chapter 2 covers some fundamental concepts on Fourier transforms, Laplace transforms, two-port network transfer functions, and the content of frequency-domain data measured using network analysers. These background definitions, terminology and nomenclature permit us to discuss matters more succinctly in later chapters. Chapter 3 introduces the idea of the phase retrieval, states some key theoretical results related to the problem in general, and provides a survey of phase retrieval methods relevant to the topic of the thesis. This allows us to nicely put the thesis topic in context. Chapter 4 contains the main contribution of the thesis. The concept of *model-based phase-retrieval* for electrical networks is introduced for the first time and the mathematical formulation discussed in detail. This is followed by application of the technique to several examples of practical interest in microwave engineering. Some general conclusions are provided in Chapter 5.

1.3 REFERENCES FOR CHAPTER 1

[1] P.A.Rizzi, "Microwave Engineering : Passive Circuits" (Prentice-Hall, 1988).

[2] *Private Communication*, COM DEV Space Group, Cambridge, Ontario, Canada.

[3] R.Levy, "Derivation of equivalent circuits of microwave structures using numerical techniques", IEEE Trans. Microwave Theory Tech., Vol.47, No.9, pp.1688-1695, Sept.1999.

[4] A.E.Atia & A.E.Williams, "Narrow-bandpass waveguide filters", IEEE Trans. Microwave Theory Tech., Vol.MTT-20, pp.258-265, April 1972.

[5] E.K.Miller & G.J.Burke, "Using model-based parameter estimation to increase the physical interpretability and numerical efficiency of computational electromagnetics", Computer Physics Communications, Vol.68, pp.43-75, 1991.

[6] X.Ding, V.K.Devabhaktuni, B.Chattaraj, M.C.E.Yagoub, M.Deo, J.Xu & Q.J.Zhang, "Neural-network approaches to electromagnetic-based modeling of passive components and their applications to high-frequency and high-speed nonlinear circuit optimization", IEEE Trans. Microwave Theory Tech., Vol.52, No.1, pp.436-449, Jan.2004

[7] P.D.T.O'Connor, *IEEE Spectrum*, July 2001, pp.18

Chapter 2

Fundamental Concepts

2.1 FUNDAMENTAL FOURIER TRANSFORM RELATIONS

2.1.1 Basic Definition

If $f(t)$ is a function of the real variable[1] t then its Fourier transform $F(\omega)$ is defined [1] by

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{-j\omega t}\, dt \tag{2.1-1}$$

The function $F(\omega)$ is in general complex, and so we can write

$$F(\omega) = \mathrm{Re}\{F(\omega)\} + j\,\mathrm{Im}\{F(\omega)\} = |F(\omega)|\, e^{j\arg\{F(\omega)\}} \tag{2.1-2}$$

The inversion expression is [1]

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\, e^{j\omega t}\, d\omega \tag{2.1-3}$$

The properties of $F(\omega)$ have been studied in great detail. Those of direct consequence to the discussion in this thesis will next be considered.

[1] It will be the time variable in this thesis, and hence is indeed real.

2.1.2 The Fourier Transform of a Real Function

If $f(t)$ is a real function, as it will be in practical circumstances since it will represent the time-domain response of a physical system, then $F(-\omega) = F^*(\omega)$, and thus $\mathrm{Re}\{F(-\omega)\} = \mathrm{Re}\{F(\omega)\}$ and $\mathrm{Im}\{F(-\omega)\} = -\mathrm{Im}\{F(\omega)\}$.

2.1.3 The Fourier Transform of a Real Causal Function

A function $f(t)$ is called causal if it is zero for negative t, that is

$$f(t) = 0 \qquad t < 0 \tag{2.1-4}$$

in which case the Fourier transform (2.1-1) becomes

$$F(\omega) = \int_0^\infty f(t)\, e^{-j\omega t}\, dt \tag{2.1-5}$$

In such cases $\mathrm{Re}\{F(\omega)\}$ and $\mathrm{Im}\{F(\omega)\}$ are not independent of each other. The one can be determined from the other through the Hilbert transform. If the causal function $f(t)$ does not have any singularities at $t = 0$ then [1,pp.200]

$$\mathrm{Im}\{F(\omega)\} = \frac{-1}{\pi} PV\left\{\int_{-\infty}^{\infty} \frac{\mathrm{Re}\{F(\xi)\}}{\omega - \xi} d\xi\right\} \tag{2.1-6}$$

$$\mathrm{Re}\{F(\omega)\} = \mathrm{Re}\{F(\infty)\} + PV\left\{\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{Im}\{F(\xi)\}}{\omega - \xi} d\xi\right\} \tag{2.1-7}$$

and are known as Hilbert transforms[2]. If $f(t)$ does not have any singularities at $t = 0$ then $F(\infty) = 0$ and hence $\mathrm{Re}\{F(\infty)\} = 0$. The notation $\mathrm{PV}\{\ldots\}$ reminds us [6,pp.740] that it is the principal value of the integral that has to be found.

The above expressions can be written in the alternative form as [1,pp.200]

$$\mathrm{Im}\{F(\omega)\} = \frac{-2\omega}{\pi} PV\left\{\int_0^{\infty} \frac{\mathrm{Re}\{F(\xi)\}}{\omega^2 - \xi^2} d\xi\right\} \tag{2.1-8}$$

$$\mathrm{Re}\{F(\omega)\} = \frac{2}{\pi} PV\left\{\int_0^{\infty} \frac{\xi\,\mathrm{Im}\{F(\xi)\}}{\omega^2 - \xi^2} d\xi\right\} \tag{2.1-9}$$

as long as $F(\omega)$ is not singular at $\omega = 0$. This follows [1,pp.200] from the fact that $\mathrm{Re}\{F(\omega)\}$ is even and $\mathrm{Im}\{F(\omega)\}$ is odd, as stated in Section 2.1.2. Note the change in the limits of the integrals. Equally importantly, we note that the notation $\mathrm{PV}\{\ldots\}$ once again reminds us that it is the principal value of the integral that has to be found. The principal value is defined in terms of a limiting process, which for (2.1-8) would be

$$\mathrm{Im}\{F(\omega)\} = \frac{-2\omega}{\pi} \lim_{\delta \to 0}\left\{\int_{-\infty}^{\omega-\delta} \frac{\mathrm{Re}\{F(\xi)\}}{\omega^2 - \xi^2} d\xi + \int_{\omega+\delta}^{\infty} \frac{\mathrm{Re}\{F(\xi)\}}{\omega^2 - \xi^2} d\xi\right\} \tag{2.1-10}$$

and is thus not necessarily convenient for computational purposes. Other forms of (2.1-8) and (2.1-9) have been derived [5,7,8] that "lead directly to the principal value" [8] without having to apply the limiting process. These read

$$\mathrm{Im}\{F(\omega)\} = \frac{2\omega}{\pi} \int_0^{\infty} \frac{\mathrm{Re}\{F(\xi)\} - \mathrm{Re}\{F(\omega)\}}{\xi^2 - \omega^2} d\xi \tag{2.1-11}$$

[2] The real and imaginary parts of the permittivity and permeability of a material satisfy the Hilbert transform relationships as well, as do various other material properties used in physics and chemistry. These are usually referred to as the Kramers-Kronig relations. In the mathematics literature the results taken as a whole are referred to as Titchmarsh's theorem [2].

$$\mathrm{Re}\{F(\omega)\} = \mathrm{Re}\{F(\infty)\} - \frac{2}{\pi}\int_0^\infty \frac{\xi\,\mathrm{Im}\{F(\xi)\} - \omega\,\mathrm{Im}\{F(\omega)\}}{\omega^2 - \xi^2} d\xi \tag{2.1-12}$$

Still another oft-quoted form of (2.1-12) is [8]

$$\mathrm{Re}\{F(\omega)\} = \mathrm{Re}\{F(0)\} - \frac{2\omega^2}{\pi}\int_0^\infty \frac{[\mathrm{Im}\{F(\xi)\}/\xi] - [\mathrm{Im}\{F(\omega)\}/\omega]}{\xi^2 - \omega^2} d\xi \tag{2.1-13}$$

If we use the substitution

$$u = \ln\left(\frac{\xi}{\omega}\right) \tag{2.1-14}$$

then some rearrangement yields the expressions (note the integration limits carefully)

$$\mathrm{Im}\{F(\omega)\} = PV\left\{\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d[\mathrm{Re}\{F(\xi)\}]}{du} \ln\left[\coth\frac{|u|}{2}\right] du\right\} = PV\left\{\frac{1}{\pi}\int_0^\infty \frac{d[\mathrm{Re}\{F(\xi)\}]}{d\xi} \ln\left|\frac{\xi+\omega}{\xi-\omega}\right| d\xi\right\} \tag{2.1-15}$$

$$\begin{aligned}\mathrm{Re}\{F(\omega)\} - \mathrm{Re}\{F(\infty)\} &= PV\left\{-\frac{1}{\pi\omega}\int_{-\infty}^{\infty} \frac{d[\xi\,\mathrm{Im}\{F(\xi)\}]}{du} \ln\left[\coth\frac{|u|}{2}\right] du\right\} \\ &= PV\left\{-\frac{1}{\pi\omega}\int_0^\infty \frac{d[\xi\,\mathrm{Im}\{F(\xi)\}]}{d\xi} \ln\left|\frac{\xi+\omega}{\xi-\omega}\right| d\xi\right\}\end{aligned} \tag{2.1-16}$$

and even

$$\mathrm{Re}\{F(\omega)\} = \mathrm{Re}\{F(0)\} - PV\left\{\frac{\omega}{\pi}\int_0^\infty \frac{d[\mathrm{Im}\{F(\xi)\}/\xi]}{d\xi} \ln\left|\frac{\xi+\omega}{\xi-\omega}\right| d\xi\right\} \tag{2.1-17}$$

It is important to observe that the above-mentioned Hilbert transform relationships apply to all causal functions. It is equally important to observe that they *do not* [3] relate $|F(\omega)|$ to $\arg\{F(\omega)\}$.

2.2 FUNDAMENTAL LAPLACE TRANSFORM RELATIONS

If $f(t)$ is a function of the real variable t, defined for $t \geq 0$, then its Laplace transform $F(s)$, if it exists, is defined by [1, pp.169]

$$F(s) = \int_0^\infty f(t)\, e^{-st}\, dt \tag{2.2-1}$$

where

$$s = \sigma + j\omega \tag{2.2-2}$$

is the complex frequency. We can write

$$F(s) = \mathrm{Re}\{F(\sigma,\omega)\} + j\,\mathrm{Im}\{F(\sigma,\omega)\} \tag{2.2-3}$$

2.3 NETWORK TRANSFER FUNCTIONS & RELATED PARAMETERS

2.3.1 The Transfer Function Concept

Although the methods to be developed in this thesis can be adapted to multi-port cases we will concentrate on two-ports. In particular we consider two-port networks consisting of *linear time-invariant passive materials*, and which have *zero initial conditions*.

---

[3] In Section 3.2.2 we will show that a Hilbert transform relationship exists between $|F(\omega)|$ and $\arg\{F(\omega)\}$, *but only if* additional constraints are imposed on $F(\omega)$.

Furthermore, we assume that at time t = 0 the network is excited by a single impressed (i.e. independent) time-varying voltage or current source, which we call the excitation. As the response we can designate either an output voltage or current.

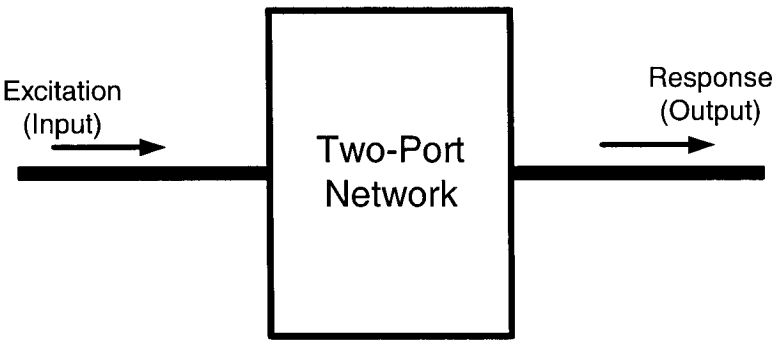


Figure 2.3-1 : System Function of a Network

Using customary notation we let E(s) represent the Laplace transform of the excitation e(t), and R(s) the Laplace transform of the response r(t). The system transfer function H(s) is then defined as

$$H(s) = \frac{R(s)}{E(s)} \qquad (2.3\text{-}1)$$

The transfer function can be voltage/voltage, voltage/current, current/current or current/voltage. As in Guillemin [3,pp.297], we will at this stage be "noncommittal" as to whether it is an impedance, admittance or dimensionless transfer function that is being referred to, so that any comments will be the same for all the above types. Of course the detailed form of the transfer function of a given network depends on the particular input-output quantities selected. It similarly stands to reason that the location of the input and output ports must be specified.

The transfer function H(s) is related to the impulse response h(t) of the network through the Laplace transform operation as [1]

$$H(s) = \int_0^\infty h(t)\, e^{-st}\, dt \qquad (2.3\text{-}2)$$

We can do this since our underlying assumption is that the system is causal and stable[4], which implies that h(t) is real, absolutely integrable, and vanishes for $t < 0$.

The transfer function $H(s)$ contains all the information on the behaviour of the system, and could be used to determine (through an inverse Laplace transform) both the transient and steady-state response of the system for a specified input. The steady-state response of the system for a single sinusoidal input of frequency ω can be determined by setting $s = j\omega$ in $H(s)$ and multiply it by the input. The function $H(j\omega)$ is called the frequency response of the system, and shows the behaviour of $H(s)$ along the ω – axis (usually called the "real frequency axis). It is clear from inspection of (2.1-4), (2.1-5) and (2.2-1) that, since the system is causal, the frequency response is just the Fourier transform of the impulse response of the system.

2.3.2 Particular Forms of Transfer Function

There are many different forms a transfer function may take. We will illustrate this through reference to Figure 2.3-2. This is a two-port network with a (total) voltage and current defined as shown at each port. The complete network is contained within the dashed-line box. The transmission lines shown are merely intended to indicate that the ports access the network through transmission lines of characteristic impedances Z_{01} and

[4] Any passive physical system (the only types with which we are concerned in this thesis) is always causal and stable.

Z_{02} that form part of the network. It is actually not necessary to mention these access lines here, but doing so will make it easier to connect the transfer functions of this section to those involving the scattering parameters in Section 2.3.3.

If the excitation (i.e. the input) is considered to be the current at Port#1 and the response (i.e. output) the voltage at Port#2, then the transfer function is a *transfer impedance*

$$H(s) = Z_{21}(s) = \left.\frac{V_2(s)}{I_1(s)}\right|_{I_2(s)=0} \tag{2.3-3}$$

If we consider $V_1(s)$ to be the excitation and $V_2(s)$ the response, then we could use the *voltage-ratio* as the transfer function

$$H(s) = \frac{V_2(s)}{V_1(s)} \tag{2.3-4}$$

Although not strictly speaking a transfer function, the driving point impedance is used in network synthesis work, is usually (albeit indirectly) one of the measured quantities in microwave networks, and is used in design work. Thus, although it is well-known, we will quickly define it here for later reference. The driving point impedance at Port#1 is

$$Z_1^{in}(s) = \frac{V_1(s)}{I_1(s)} \tag{2.3-5}$$

with a similar definition for $Z_2^{in}(s)$ at Port#2.

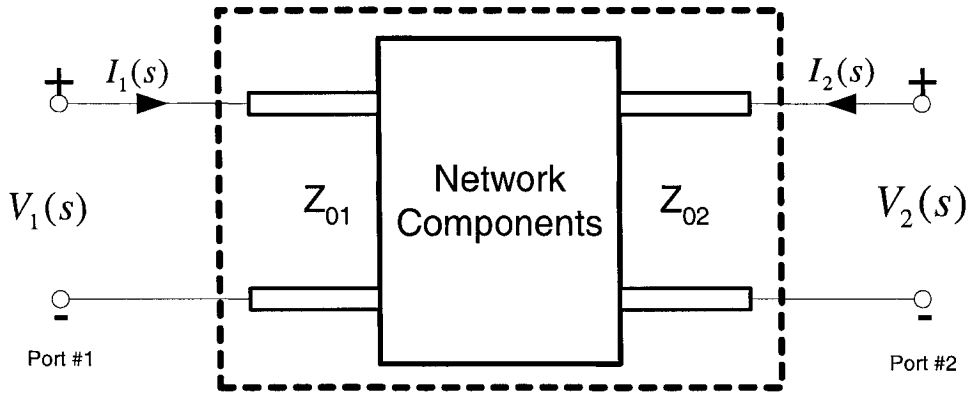


Figure 2.3-2 : Two-Port Network

2.3.3 Scattering Parameters as Transfer Functions

The two-port microwave network shown in Figure 2.3-3 can be characterized in terms of four scattering parameters $S_{11}(\omega)$, $S_{12}(\omega)$, $S_{21}(\omega)$ and $S_{22}(\omega)$, where $\omega = 2\pi f$ is the angular frequency and f is the frequency. The scattering parameters $S_{ij}(\omega)$ are complex quantities and thus have a magnitude $\left|S_{ij}(\omega)\right|$ and phase $\arg\left\{S_{ij}(\omega)\right\}$.

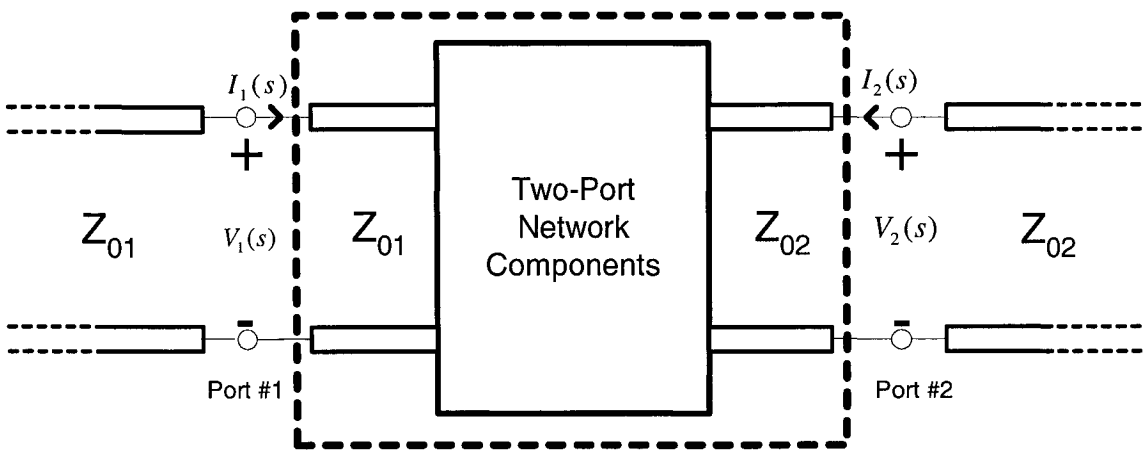


Figure 2.3-3 : Two-Port Network Emphasising Transmission Lines of the Same Characteristics Impedances as the Access Lines Connected at Each End

In the situation shown in Figure 2.3-4, consider that we have the generator impedance Z_g = Z_{01} and the load impedance $Z_L = Z_{02}$. These are the circumstances that are closely approximated when making network analyzer measurements. In such cases we then [5] have[5]

$$S_{21}(\omega) = \frac{V_2}{V_g}\sqrt{\frac{Z_{01}}{Z_{02}}} \tag{2.3-6}$$

and thus $S_{21}(\omega)$ has the dimensions of a voltage-ratio transfer function, and so we can refer to it as the transfer function of the two-port network. Similar comments can be made for parameter $S_{12}(\omega)$ if the generator and load are swapped around. In fact, very many circumstances (we could probably even say most circumstances) are such that $Z_{01} = Z_{02} = Z_0$ and then

$$S_{21}(\omega) = \frac{V_2}{V_g} \tag{2.3-7}$$

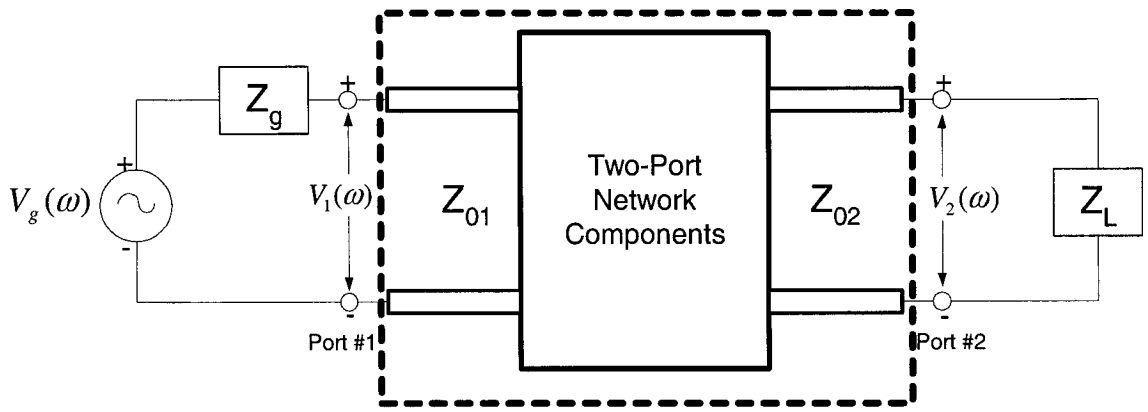


Figure 2.3-4 : Two-Port Network with a Generator and Load

[5] Recall that V_2 is the *total* voltage at Port#1.

Observe that the parameter $S_{11}(\omega)$ is related to the driving point impedance through

$$S_{11}(\omega) = \frac{Z_1^{in}(\omega) - Z_{01}}{Z_1^{in}(\omega) + Z_{01}} \tag{2.3-8}$$

and similarly for $S_{22}(\omega)$.

2.3.4 Related Parameters

In Section 2.3.3 we said that we can write $S_{21}(\omega) = |S_{21}(\omega)|e^{j\arg\{S_{21}(\omega)\}}$, and will now for notational convenience write the phase response using the symbol

$$\phi(\omega) = \arg\{S_{21}(\omega)\} \tag{2.3-9}$$

A parameter known as the group delay, which is [9,pp.9]

$$\tau_g(\omega) = -\frac{\partial\phi(\omega)}{\partial\omega} \tag{2.3-10}$$

is an important performance measure in many systems since it has deleterious effects on certain modulation schemes. Since $\omega = 2\pi f$, it is important to remember that if (as is usually the case when working with measured data) the phase is available as a function of f rather than ω, then

$$\tau_g(f) = -\frac{1}{2\pi}\frac{\partial\phi(f)}{\partial f} \tag{2.3-11}$$

in nano-seconds when $\phi(f)$ is in radians and f in GHz, and

$$\tau_g(f) = -\frac{1}{360}\frac{\partial\phi(f)}{\partial f} \tag{2.3-12}$$

in nano-seconds when $\phi(f)$ is in degrees and f in GHz.

Also significant in many instances is the group delay variation, which is measured in one of two ways. Firstly, over a given frequency band $\omega_L \le \omega \le \omega_U$, we will find that $\tau_g(\omega)$ varies between an upper bound $\tau_g^{\max}$ and a lower bound $\tau_g^{\max}$. The difference $\left|\tau_g^{\max} - \tau_g^{\min}\right|$ is called the *group delay ripple* over the operating frequency band of interest. Secondly, the value of the derivative of the group delay with respect to frequency, in other words

$$\tau_g^{slope} = \frac{\partial \tau_g(\omega)}{\partial \omega} \tag{2.3-13}$$

is called the group delay slope. It is usually the maximum value of the group delay slope

$$\tau_g^{slope} = \max\left\{\frac{\partial \tau_g(\omega)}{\partial \omega}\right\}_{\omega_L \le \omega \le \omega_U} \tag{2.3-14}$$

that is of interest. Sometimes system specifications place upper bounds on these values.

2.4 NETWORK TRANSFER FUNCTION PROPERTIES

2.4.1 Rational Function Form of Transfer Functions

If a network can be fully described by ordinary differential equations (eg. if it is comprised of solely lumped components) then its transfer function can be expressed as a rational function in terms of a finite number of poles and zeros. We are able to write

$$H(s) = \frac{N(s)}{D(s)} \tag{2.4-1}$$

where $N(s)$ and $D(s)$ are polynomials. The zeros of the numerator polynomial $N(s)$ are the zeros of the transfer function. The zeros of the denominator polynomial $D(s)$ are the poles of the transfer function. Any passive physical system is stable, and its poles are all situated on the left hand side of the s-plane[6]. Its zeros may be situated anywhere. More will be said about these in Section 2.4.2. It can also be shown that any zeros or poles which are not real must occur in complex conjugate pairs.

If a network can only be completely described by a partial differential equation (eg. if it explicitly contains distributed components) then the number of poles and zeros is infinite. As we might expect, most microwave circuits thus possess an infinite number of poles and/or zeros. If we wish to use a model containing poles and zeros for such networks, then in order for the model to be computationally feasible, it is necessary to approximate this infinite set of poles and zeros by performing computations using only the dominant poles and zeros of the network over the frequency band $\omega_L \leq \omega \leq \omega_U$ of

[6] Strictly speaking, a stable system can have poles situated on the real frequency axis, but actual physical systems, which always have some loss, will not.

interest . In other words, we use an approximate representation of the transfer function usually referred to as a reduced order model.

2.4.2 Minimum Phase and Non-Minimum Phase Transfer Functions

A network which has zeros in the left-half plane or on the ω axis only, is called a *minimum phase* network. If its transfer function has one or more zeros in the right-half plane it is called a *non-minimum phase* network. A maximum-phase network is one which has all its transfer function zeros in the right-half plane. It is worthwhile to examine the reason for the above terminology. We therefore consider two networks whose pole-zero diagrams are shown in Figure 2.4-1. The networks have the same poles, but the zeros are mirror images of each other. According to the above classification the network described by Figure 2.4-1(a) is a minimum phase network while that described by Figure 2.4-1(b) is a non-minimum phase network. If we were to plot the magnitude of the transfer functions in each case we would find that they are identical. However, the phase responses are those plotted in Figure 2.4-2. It is clear that the absolute value of the phase is larger for the non-minimum phase network than for the minimum phase network at all frequencies; hence the terminology used.

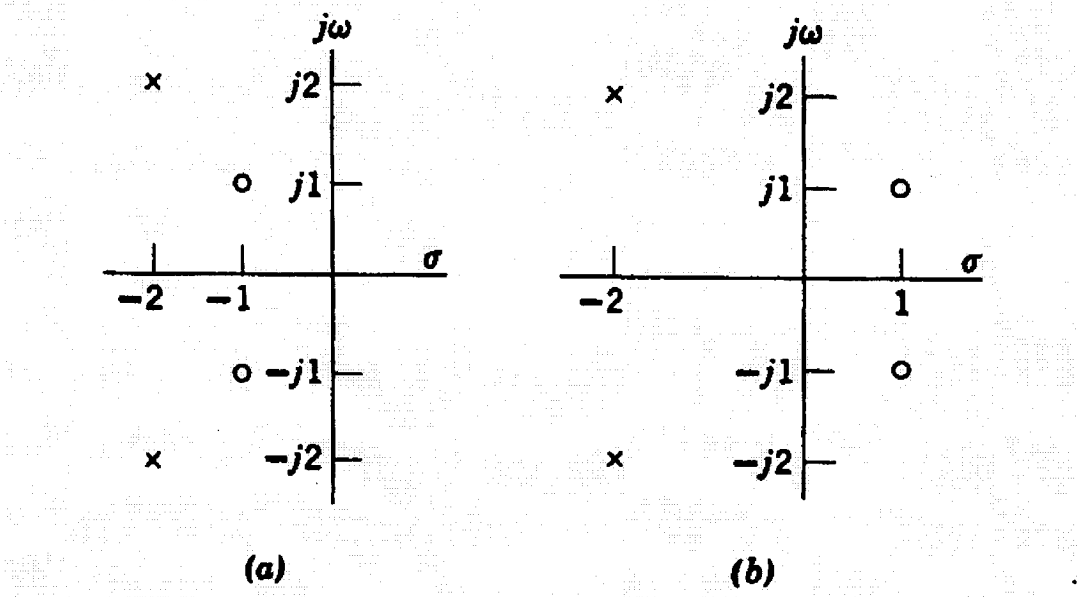


Figure 2.4-1 : Pole-Zero Plot of (a). Minimum-Phase Function and (b). Non-Minimum Phase Function (After [6]).

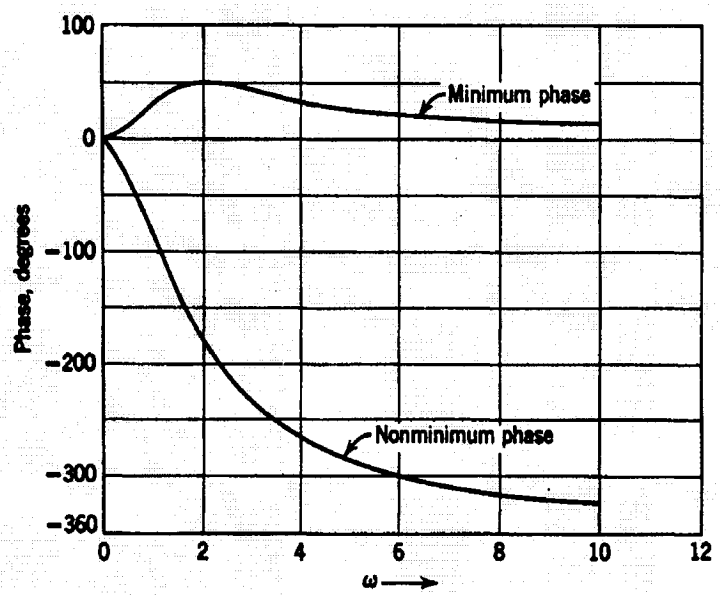


Figure 2.4-2 : Phase Responses of the Networks with Pole-Zero Distributions Shown in Fig.2.4-1 (After [6]).

2.4.3 Non-Minimum Phase Transfer Functions Expressed in Terms of Minimum-Phase and All-Pass Transfer Functions

Although it is not a central idea to the techniques developed in this thesis we will need to mention the so-called all-pass network in Chapter 3, and so we introduce it here. The all-pass network is a non-minimum phase network with a very specific response. Consider a class of networks whose poles (as with all stable networks) are only in the left half-plane, and whose zeros are mirror images of each of these poles about the $j\omega$. Thus a network with a transfer function

$$H_{ALL}(s) = \frac{(s - \alpha_1 - j\omega_1)}{(s + \alpha_1 - j\omega_1)} \tag{2.4-2}$$

is an example in this class. For the above transfer function it is clear that (by setting $s = j\omega$) we have

$$H_{ALL}(\omega) = \frac{(j\omega - \alpha_1 - j\omega_1)}{(j\omega + \alpha_1 - j\omega_1)} = \frac{\alpha_1 + j(\omega - \omega_1)}{-\alpha_1 + j(\omega - \omega_1)} \tag{2.4-3}$$

and so

$$|H_{ALL}(\omega)| = \left|\frac{\alpha_1 + j(\omega - \omega_1)}{-\alpha_1 + j(\omega - \omega_1)}\right| = \frac{\sqrt{\alpha_1^2 + (\omega - \omega_1)^2}}{\sqrt{\alpha_1^2 + (\omega - \omega_1)^2}} = 1 \tag{2.4-4}$$

In other words, the transfer function magnitude is independent of frequency. Hence the name "all-pass network". The phase function of the network is a function of frequency. It is for the above reasons that all-pass networks are used as phase equalisers in order to correct for phase distortions caused by other networks in a system comprised of a cascade of several networks. The above considerations can be generalised by recognising that all-pass transfer functions can be written as

$$H_{ALL}(s) = \frac{\prod_{n=1}^{N}(s - \alpha_n - j\omega_n)}{\prod_{n=1}^{N}(s + \alpha_n - j\omega_n)} \tag{2.4-5}$$

where N depends on the order of the network.

A more specific example of an all-pass network pole-zero layout is given in Figure 2.4-3, and the phase and magnitude response are provided in Figure 2.4-4. As expected, the magnitude of the response is the same for all frequencies but the phase response is not. It is important to remember that all-pass networks are non-minimum phase networks.

If we have a particular arbitrary transfer function $G(\omega)$, then all transfer functions $G(\omega)H_{ALL}(\omega)$ formed by the multiplication of $G(\omega)$ and the all-pass transfer function will have the same magnitude response since $|G(\omega)H_{ALL}(\omega)| = |G(\omega)||H_{ALL}(\omega)| = |G(\omega)|$. It is for this reason that, given a network with a transfer function $H(s)$, it is always possible to write it in the form

$$H(\omega) = H_{MIN}(\omega)H_{ALL}(\omega) \tag{2.4-6}$$

where

- $H_{MIN}(\omega)$ is the transfer function of a minimum phase network.
- $H_{ALL}(\omega)$ is the transfer function of an all-pass network, so that $|H_{ALL}(\omega)| = 1$.

This result is utilised in certain phase retrieval algorithms reviewed in Chapter 3.

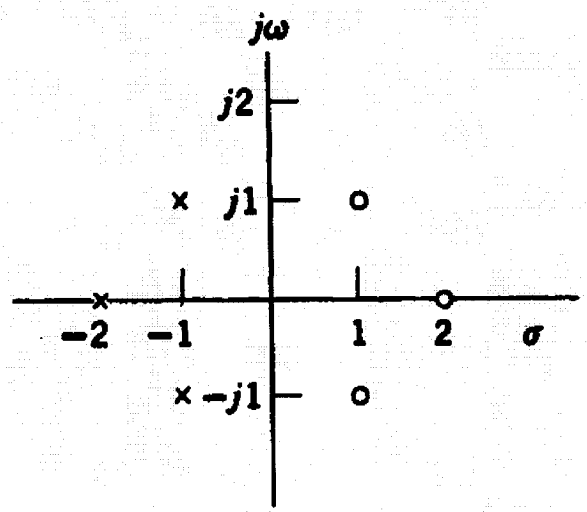


Figure 2.4-3 : Pole-Zero Plot of a Specific All-Pass Network (After [6]).

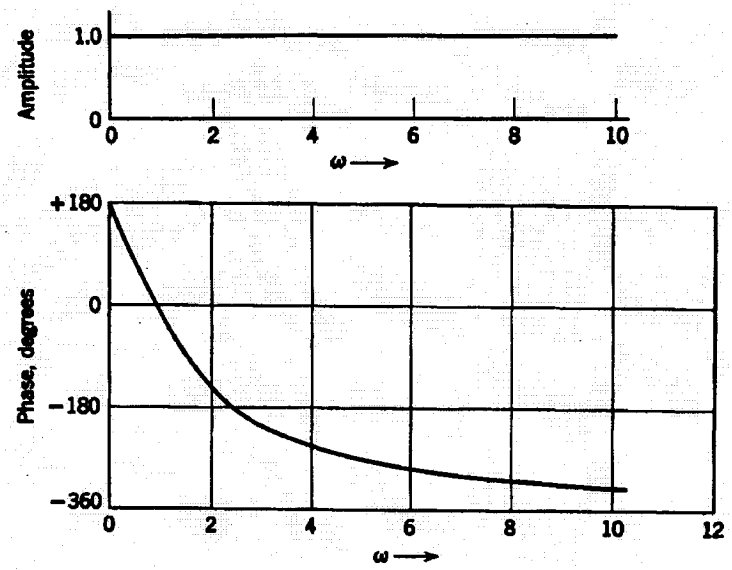


Figure 2.4-4 : Magnitude and Phase Responses of the All-Pass Networks with Pole-Zero Distributions Shown in Fig.2.4-1 (After [6]).

2.4.4 Examples of Minimum and Non-Minimum Phase Circuits

An intuitive appreciation of the minimum and non-minimum phase network concepts can be gained by providing some specific circuit examples. It is noted by Guillemin [3,pp.47] that any physically realisable passive ladder network like that shown in Figure 2.4-5 always has a minimum phase transfer impedance.

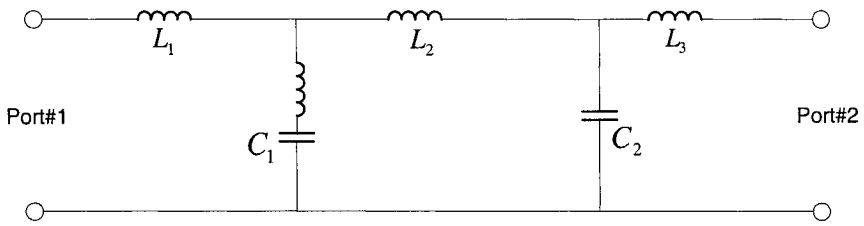


Figure 2.4-5 : Two-Port Ladder Network as an Example of a Minimum-Phase Network

Bode [6,pp.243] remarks that "circuits which are broadly not of the ladder type are those in which the current can reach the load by alternative paths" and then provides two specific examples, slightly edited versions of these are shown here in Figure 2.4-6 and 2.4-7.

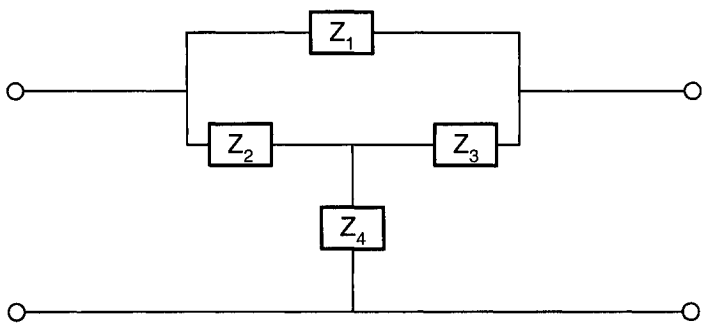


Figure 2.4-6 : First Specific Example of a Non-Minimum Phase Two-Port Network

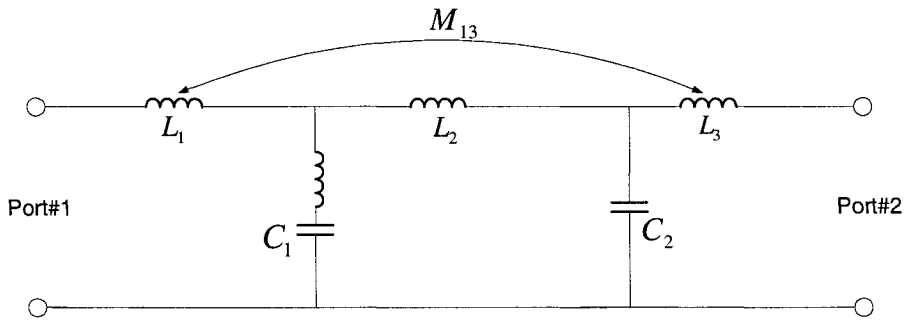


Figure 2.4-7 : Second Specific Example of a Non-Minimum Phase Two-Port Network

The network in Figure 2.4-7 is almost identical to that in Figure 2.4-5. The only difference is the mutual coupling between components L_1 and L_3. Re-drawing this as shown in Figure 2.4-8 reveals more clearly the "alternative current path" referred to by Bode [6,pp.243].

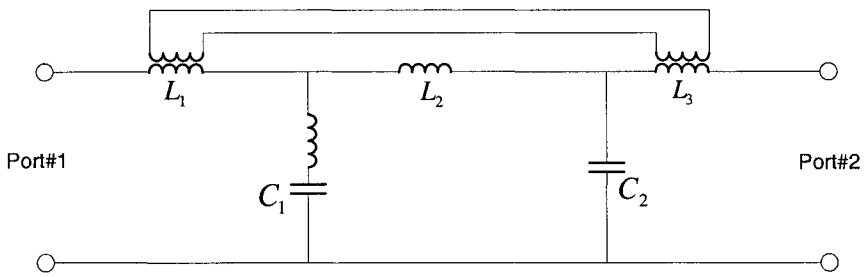


Figure 2.4-8 : Redrawn Version of the Non-Minimum Phase Two-Port Network of Figure 2.4-3

2.5 THE CONTENT OF MEASURED FREQUENCY-DOMAIN DATA

The vector network analyzer provides the scattering parameters $S_{ij}(\omega)$ of the two-port network under test. In the interests of brevity, we will in this section denote any of the S-

parameters by $H(\omega)$. The measurement does not supply the quantity $H(\omega)$ over a continuum of frequencies ω, or over the entire infinite frequency range $-\infty \leq \omega \leq \infty$. Instead, the instrument measures each of the scattering parameters at the discrete frequencies $\{\omega_0, \omega_1, \omega_2, ..., \omega_n, ..., \omega_{N-1}\}$, namely the set

$$\{H(\omega_0), H(\omega_1), H(\omega_2), ..., H(\omega_n), ..., H(\omega_{N-1})\} \tag{2.5-1}$$

where $\omega_0 = \omega_L$ and $\omega_{N-1} = \omega_U$, the subscripts "L" and "U" denoting "lower" and "upper", respectively. In most situations of practical interest this does not imply that $H(\omega)$ is bandlimited (i.e. it does not imply that $H(\omega) = 0$ for $\omega < \omega_L$ or $\omega > \omega_L$). All it says is that we do not know $H(\omega)$ except in the frequency range $\omega_L \leq \omega \leq \omega_U$. Indeed, we know that $H(\omega)$ will unlikely be bandlimited. The vector network analyzer calculates phase-related quantities such as the group delay by numerically differentiating the phase of S_{21} and S_{12}, followed by some smoothing.

In the case of scalar measurements, all that is available is

$$\{|H(\omega_0)|, |H(\omega_1)|, |H(\omega_2)|, ..., |H(\omega_n)|, ..., |H(\omega_{N-1})|\} \tag{2.5-2}$$

It is this frequency-sampled version of $|H(\omega)|$, available over a limited range of frequencies, that we have to work with in the phase-retrieval considerations in this thesis.

The question now is this : If we wish to consider $H(\omega)$ to be a Fourier transform, what is its inverse transform? We know from Section 2.1-1 that in theory it is

$$h(t) = \int_{-\infty}^{\infty} H(\omega) e^{j\omega t} \, d\omega \tag{2.5-3}$$

We in effect have (before sampling)

$$H_{MEASURED}(\omega) = H(\omega)H_{BPF}(\omega) \tag{2.5-4}$$

where the ideal bandpass transfer function $H_{BPF}(\omega)$ is

$$H_{BPF}(\omega) = \begin{cases} 1 & \omega_L \le \omega \le \omega_U \\ 0 & Otherwise \end{cases} \tag{2.5-5}$$

This fact will have to be faced when dealing with any phase-retrieval problem.

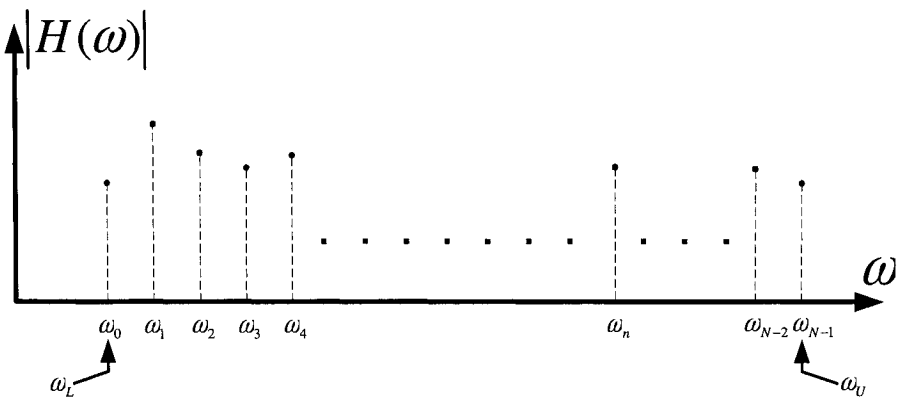


Figure 2.5-1 : Illustration of the Sampled Values of $|H(\omega)|$

2.6 CONCLUDING REMARKS

In this chapter we have collected together, and clarified, some definitions and theoretical ideas that we will need to refer to when discussing phase-retrieval in the chapters of the thesis that follow.

2.7 REFERENCES FOR CHAPTER 2

[1] A.Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, 1962).

[2] H.M.Nussenzveig, *Causality and Dispersion Relations* (Academic Press, 1972).

[3] E.A.Guillemin, *Synthesis of Passive Networks* (Wiley, 1957)

[4] P.A.Rizzi, "Microwave Engineering : Passive Circuits" (Prentice-Hall, 1988).

[5] H.W.Bode, *Network Analysis and Feedback Design* (Van Nostrand Co., 1955).

[6] F.F.Kuo, *Network Analysis and Synthesis* (Wiley, 1966).

[7] A.T.Starr, *Radio and Radar Technique* (Pitman & Sons Ltd., 1953).

[8] J.R.James & G.Andrasic, "Assessing the accuracy of wideband electrical data using Hilbert transforms", IEE Proc., Vol.137, Pt.H, No.1, pp.184-188, June 1990.

[9] J.D.Rhodes, *Theory of Electrical Filters* (Wiley, 1976).

Chapter 3

Existing Phase Retrieval Algorithms

3.1 THE PHASE RETRIEVAL PROBLEM DEFINED

The formal statement of the one-dimensional phase retrieval problem is quite straightforward :

> Given the magnitude $|H(\omega)|$ of a transfer function $H(\omega)$ over $\omega_L \leq \omega \leq \omega_U$,
> determine $\arg\{H(\omega)\}$ over $\omega_L \leq \omega \leq \omega_U$.

The terms "phase recovery" and "signal reconstruction" are alternative names used for the phase retrieval problem in different applications contexts, although these usually refer to two-dimensional problems.

The measured data from many physical systems (eg. astronomy, optical systems, microwave systems) can be viewed as possessing magnitude[7] (or intensity, which is the square of the magnitude) and phase. In many cases it is not possible, or is difficult, or is excessively time-consuming, to measure phase. As a consequence, the phase-retrieval problem arises in many branches of physics and engineering [1], and thus copious references exist on the topic. The reasons given above that a full vector (i.e. magnitude and phase) measurement might not be possible because it is simply too time-consuming

[7] We note that in the phase retrieval literature the term "amplitude" is usually taken to mean "magnitude plus the sign of the phase".

might appear absurd from the point of view of those engaged in measurements that are being done for fundamental scientific value, but is not necessarily so in engineering applications.

In the present one-dimensional case, where $H(\omega)$ is a function of a single variable ω, it can be shown mathematically [1] that given $|H(\omega)|$ there is no unique phase $\arg\{H(\omega)\}$ associated with $H(\omega)$. In other words, it is not possible to retrieve a unique phase if only $|H(\omega)|$ has been specified. It is *always necessary* to specify additional constraints (i.e. more information about the transfer function) in order to retrieve $\arg\{H(\omega)\}$ from a given $|H(\omega)|$. This reality is not always appreciated, possibly because of the fact that for m-dimensional Fourier transforms (m > 1), there is almost always a unique relationship between the Fourier transform magnitude and its phase. The one can in principle be recovered from the other in a unique manner. As explained by Bates and McDonnell [2,pp.110], the terminology *almost always* implies that "solutions to multidimensional Fourier phase (retrieval) problems are unique except in cases so special that they have no practical significance". Of course, knowing that this is so does not tell us how to actually determine $\arg\{H(\omega)\}$ once we have $|H(\omega)|$. Thus there are many different phase retrieval algorithms, the reason for the existence of many different ones being that some are more successful than others.

Due to the above-mentioned non-uniqueness property possessed by one-dimensional problems not all of the phase retrieval methods that have been developed for two-dimensional problems are directly relevant for the problem of interest in the present thesis. Thus we will not attempt a survey of the entire field. Instead only those that could

possibly be applied to the one-dimensional case, or that have a bearing on some aspect we wish to emphasise later in the thesis, will be considered.

3.2 ROUTES FOLLOWED BY EXISTING ONE-DIMENSIONAL PHASE RETRIEVAL ALGORITHMS

We mentioned in Section 3.1 that in the one-dimensional case of interest here a given complex transfer function $H(\omega)$ has associated with it a unique phase $\arg\{H(\omega)\}$, but that there is no unique $\arg\{H(\omega)\}$ associated with a specified magnitude $|H(\omega)|$ of a transfer function. In fact there are an infinite number of possibilities for the $\arg\{H(\omega)\}$ consistent with a particular $|H(\omega)|$. In order to derive a unique phase response from the given magnitude response more information about the transfer function must be assumed. In existing phase retrieval algorithms these include :

(1). Utilisation of a minimum-phase assumption.

(2). Utilisation of the values of $\arg\{H(\omega)\}$ at two frequency points ω_a and ω_b, namely $\arg\{H(\omega_a)\}$ and $\arg\{H(\omega_b)\}$.

(3). Utilisation of some time-samples of $h(t)$.

(4). Utilisation of an assumption that $H(\omega)$ is bandlimited, or that $h(t)$ is time-limited.

3.3 UTILISATION OF THE MINIMUM-PHASE CONSTRAINT

3.3.1 Fundamental Relations

In the case of minimum phase networks there is a unique relationship between the magnitude response and the phase response at real frequencies. This is best dealt with by rewriting the transfer function in the form

$$H(\omega) = e^{-g(\omega)} e^{-j\phi(\omega)} \tag{3.3-1}$$

We then have

$$\phi(\omega) = \arg\{H(\omega)\} \tag{3.3-2}$$

and

$$|H(\omega)| = e^{-g(\omega)} \tag{3.3-3}$$

or in other words

$$g(\omega) = -\ln|H(\omega)| \tag{3.3-4}$$

Thus (3.3-1) can also be written as

$$H(\omega) = e^{\ln|H(\omega)|} e^{-j\arg\{H(\omega)\}} \tag{3.3-5}$$

Then, if $H(\omega)$ is a minimum-phase function, $g(\omega)$ and $\phi(\omega)$ are related through the Hilbert transform as [3,pp.206]

$$\phi(\omega) = \frac{\omega}{\pi}\int_{-\infty}^{\infty}\frac{g(\xi)}{\xi^2-\omega^2}d\xi = -\frac{\omega}{\pi}\int_{-\infty}^{\infty}\frac{\ln|H(\xi)|}{\xi^2-\omega^2}d\xi = -\frac{2\omega}{\pi}\int_{0}^{\infty}\frac{\ln|H(\xi)|}{\xi^2-\omega^2}d\xi \tag{3.3-6}$$

$$g(\omega) = g(0) - \frac{\omega^2}{\pi}\int_{-\infty}^{\infty}\frac{\phi(\xi)}{\xi(\xi^2-\omega^2)}d\xi \tag{3.3-7}$$

and so $\phi(\omega)$ can be uniquely determined from $g(\omega)$. Similarly $g(\omega)$ can be uniquely determined from $\phi(\omega)$ and the value $g(0)$. This uniqueness has come about due to the

assumption that $H(\omega)$ is a minimum-phase function. If it is not, then the uniqueness no longer holds. There are other forms [4,5,6,7] of the expressions relating $g(\omega)$ and $\phi(\omega)$. One form used by Bode [4,pp.313-315] provides a particularly intuitive understanding of the expressions. However, since these will not actually be used in this thesis the reader is referred to the above-mentioned references.

3.3.2 Use of the Minimum-Phase Case in Practice

Phase retrieval based on the minimum-phase assumption and the resulting Hilbert transform relationships has been described by Perry and Brazil [8,9,10]. These authors define what they call the *Hilbert-transform derived relative group delay* (HGD) which they show is approximately the group delay of the two-port network minus a constant. Perry and Brazil [9] observe about their technique that "if the principal structure of the transfer function (say the pass band of a filter) is contained within the measurement band ... then the difference between the actual group delay of the circuit under test and the HGD as calculated above will be approximately a constant across the majority of the measurement band". In many instances, such as those that will be dealt with in Chapter 4, it is uncertain how to define the "principal structure of the transfer function". Consider a length of coaxial cable with a connector at each end, as will be discussed in Section 4.5. If we decide we wish to use it, and hence test it, over some limited frequency band, we know that it could be used over a much wider band and so the band over which it is tested will not contain the entire "principal structure of the transfer function". Thus the minimum-phase approach is not necessarily applicable.

3.4 UTILISATION OF ALTERNATIVE CONSTRAINTS

Other existing phase-retrieval algorithms (eg. [11] through [38]) need one or two phase points in addition to magnitude everywhere. Alternatively, information at one or more data points in the time-domain version of the data is required. None of these are acceptable for the class of problems of interest in this thesis. Nevertheless, a brief review is appropriate.

In the digital signal processing literature the aim is usually to reconstruct (or equivalently perform phase retrieval on) a signal, say $h(t)$, from the magnitude $|H(\omega)|$ of its Fourier transform. These are not the transfer functions of some system, and one has some control over the time-interval or frequency-interval over which these functions are significant. Thus reference in this area usually make the assumption that the functions involved are both band-limited and time-limited. In [15] the signal reconstruction algorithm (equivalently the phase-retrieval algorithm) requires that one know the sign of the phase of $|H(\omega)|$, and that $|H(\omega)|$ must be non-zero over the frequency band of interest. But in the present application this defeats the whole purpose of just wanting to measure $|H(\omega)|$. In the portion [13] that deals with signal reconstruction from the magnitude of the Fourier transform, some rather complicated restrictions are required that exclude most situations of practical interest, as the authors themselves admit. In [12,16,28] signal reconstruction from $|H(\omega)|$ is done but some time samples of $h(t)$ are also required. The method of [33,34] similarly requires some time-samples of $h(t)$ along with $|H(\omega)|$; it determines the phase from $|H(\omega)|$ by initially assuming the system to be minimum phase, and then utilizes the time-samples of $h(t)$ to arrive at the non-minimum

phase solution from the minimum phase solution, in essence using the result (2.4-6). In [25] phase retrieval is possible from $|H(\omega)|$ and a knowledge of the initial time-sample of $h(t)$. However, if we knew $h(t)$ then we could find $\arg\{|H(\omega)|\}$ anyway. Thus while this approach may have its uses in some scenarios it is not useful for the situations of interest in this thesis.

Reference [20, 23] considers the $|H(\omega)|$ to be specified over a finite frequency segment, assumes it is periodic outside this segment, and thus assumes $H(\omega)$ can be represented as a Fourier series (in terms of sines and cosines, each component having an amplitude and a phase). Using the known relationship between $\mathrm{Re}\{H(\omega)\}$ and $\mathrm{Im}\{H(\omega)\}$ for causal systems, the authors of [23] are able to formulate a phase-retrieval algorithm as a minimization of the difference between the measured $|H(\omega)|$ and its series representation. The series coefficients, namely the component amplitudes and phases, are the variables that are altered during the optimisation process. Once these are known they can be re-substituted into the series representation to find $H(\omega)$ and hence $\arg\{|H(\omega)|\}$. However, the actual and retrieved phase differ by a linear function of frequency. This is not a problem if only the derivative of the phase response is of interest (eg. for group delay determination) and the input/output signals travel in a non-dispersive medium. In many microwave applications neither of the above restrictions are acceptable. Reference [24] is similar to [23], except that it describes a computationally more efficient form of the algorithm in [23] that is applicable when $|H(\omega)|$ is symmetrical. The method in [22] is similar to that in [23,24], except that is uses the complex Fourier series (to allow use of

discrete Fourier transforms). Its limitations are the same as those mentioned for [23]. These same limitations appear in the method of [26].

In [29, 30] the phase-retrieval algorithm uses $|H(\omega)|$ but also requires at least two values of the complex $H(\omega)$, at say ω_a and ω_b. In [35,36] the signal $h(t)$ is reconstructed (and hence its phase $\arg\{|H(\omega)|\}$ retrieved) using $|H(\omega)|$ and the Fourier transform magnitude of a signal obtained by adding a known reference signal to $h(t)$. This could possibly be used in the microwave engineering context but would require the insertion of additional components in the test set-up. A similar approach was in fact described in [17], although not referenced in [38]. The method in [19] retrieves $\arg\{|H(\omega)|\}$ from $|H(\omega)|$, but only if $|h(t)|$ is also known. Once more, Thus while having the above-mentioned additional data may have its uses in some measurement situations, it is not useful for those of interest in this thesis.

Two-dimensional phase retrieval [31, 37-45] has regularly been discussed in the literature in relation to antenna near-zone testing applications, although not yet widely used in practice. While it remains true that we are in this thesis solely concerned with one-dimensional data, we thought it appropriate to mention these applications since they might be more familiar in the microwave engineering community than those being discussed in the present thesis. Unfortunately, because of the fact that the magnitude-phase relationship is "almost always" unique for m-dimensional problems (m > 1) the methods used in the above references do not provide us any relief for one-dimensional ones. They essentially assume uniqueness, and then discuss algorithms in terms of how rapidly (if at all) they converge to the correct solution for the retrieved phase.

3.5 CONCLUSIONS

This chapter has formally defined what is meant by the term "phase retrieval problem". It has briefly reviewed some existing phase retrieval algorithms and indicated why these are not directly applicable to the problems of interest in this thesis. We wish to have a phase-retrieval method that will work for both minimum-phase and non-minimum-phase networks, and which requires the measured magnitude of the frequency response and no other measurements.

3.6 REFERENCES FOR CHAPTER 3

[1] L.S.Taylor, "The phase retrieval problem", IEEE Trans. Antennas & Propagation, Vol.AP-29, No.2, pp.386-391, March 1981.

[2] R.H.T.Bates & M.J.McDonnell, *Image Restoration and Reconstruction* (Oxford University Press, 1989).

[3] A.Papoulis, *The Fourier Integral and its Applications* (McGraw-Hill, 1962).

[4] H.W.Bode, *Network Analysis and Feedback Design* (Van Nostrand Co., 1955).

[5] E.A.Guillemin, *Synthesis of Passive Networks* (Wiley, 1957)

[6] A.T.Starr, *Radio and Radar Technique* (Pitman & Sons Ltd., 1953).

[7] J.R.James & G.Andrasic, "Assessing the accuracy of wideband electrical data using Hilbert transforms", IEE Proc., Vol.137, Pt.H, No.1, pp.184-188, June 1990.

[8] P.Perry & T.J.Brazil, "Estimation of group delay ripple of physical networks using a fast, bandlimited Hilbert transform algorithm and scalar transfer function measurements", Proc. 25th European Microwave Conference, pp.1259-1264, 1995.

[9] P.Perry & T.J.Brazil, "Hilbert-transform-derived relative group delay measurement of frequency conversion systems", IEEE International Microwave Symposium Digest, pp,1695-1698, 1996.

[10] P.Perry & T.J.Brazil, "Hilbert-transform-derived relative group delay", IEEE Trans. Microwave Theory & Techniques, Vol.45, No.8, pp,1204-1225, Aug.1997.

[11] A.V.Oppenheim & R.W.Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, 1989).

[12] A.E.Yagle & A.E.Bell, "One- and two-dimensional minimum and nonminimum phase retrieval by solving linear systems of equations", IEEE Trans. Signal Processing, Vol.47, No.11, pp.2978-2989, Nov.1999.

[13] M.Hayes & A.V.Oppenheim, "Signal reconstruction from phase or magnitude", IEEE Trans. Acoustics, Speech & Signal Processing, Vol.ASSP-28, No.6, pp.672-680, Dec.1980.

[14] M.L.Liou & C.F.Kurth, "Computation of group delay from attenuation characteristics via Hilbert transformation and spline function and its application to design", IEEE Trans. Circuits Systems, Vol.CAS-22, No.9, pp.729-734, Sept.1975.

[15] P.L. van Hoye, M.H.Hayes, J.S.Lim & A.V.Oppenheim, "Signal reconstruction from signed Fourier transform magnitude", IEEE Trans. Acoustics, Speech & Signal Processing, Vol.ASSP-31, No.5, pp.1286, Oct.1983.

[16] A.E.Yagle, "Phase retrieval from Fourier magnitude and several initial time samples using Newton's formulae", IEEE Trans. Signal Processing, Vol.46, No.7, pp.2054-2056, July 1998.

[17] J.G.Walker, "The phase retrieval problem : A solution based on zero location by exponential apodisation", Optica Acta, Vol.28, No.6, pp.735-738, 1981.

[18] M.R.Teague, "Deterministic phase retrieval : A Green's function solution", J.Opt.Soc.Am., Vol.73, No.11, pp.1434-1441, Nov.1983.

[19] R.A.Gonsalves, "Phase retrieval from modulus data", J.Opt.Soc.Am., Vol.66, No.9, pp.961-964, Sept.1976.

[20] J.Yang & T.K.Sarkar, "Reconstructing of a non-minimum phase response from far-field power pattern of an electromagnetic system", Proc. 20th Annual Review of Progesss in Applied Computational Electromagnetics, April 2004.

[21] C.Rusu & P.Kuosmanen, "Phase approximation by logarithmic sampling of gain", IEEE Trans. Circuits & Systems – II : Analog and Digital Signal Processing, Vol.50, No.2, pp.93-101, Feb.2003.

[22] J.Yang, J.Koh & T.K.Sarkar, "Reconstructing a nonminimum phase response from the far-field power pattern of an electromagnetic system", IEEE Trans. Antennas & Propagation, Vol.53, No.2, pp.833-841, Feb.2005.

[23] T.K.Sarkar & B.Hu, "Generation of nonminimum phase from amplitude-only data", IEEE Trans. Microwave Theory & Techniques, Vol.46, No.8, pp,1079-1084, Aug.1998.

[24] J.Koh, Y.Cho & T.K.Sarkar, "Reconstruction of non-minimum phase function from only amplitude data", Microwave & Optical Technology Letters, Vol.35, No.3, pp.212-216, Nov.2002

[25] H.Sahinoglou & S.D.Cabrera, "On phase retrieval of finite-length sequences using the initial time sample", IEEE Trans. Circuits & Systems, Vol.38, No.8, pp.954-958, Aug.1991.

[26] E.Brinkmeyer, "Simple algorithm for reconstructing fiber gratings from reflectometric data", Optics Letters, Vol.20, No.8, pp810-812, April 1995.

[27] A.W.Attiya, "Three transmission-line transformers for phase retrieval from scalar reflection coefficients", Microwave & Optical Technology Letters, Vol.40, No.3, pp.231-235, Feb.2004.

[28] A.E.Bell & A.E.Yagle, "Discrete phase retrieval by solving linear systems of equations : Performance under noisy conditions", Proc.ICASSP, pp.717-721, 1998.

[29] E.M.Vartiainen, K.-E.Peiponen, H.Kishida & T.Koda, "Phase retrieval in nonlinear optical spectroscopy by the maximum-entropy method : An application to the $\left|\chi^{(3)}\right|$ spectrum", J.Opt.Soc.Am.B, Vol.13, No.10, pp2106-2114, Oct.1996.

[30] J.Ahola, E.M.Vartiainen & T.Lindh, "Phase retrieval from impedance amplitude measurement", IEEE Power Electronics Letters, Vol.3, No.2, pp.50-52, June 2005.

[31] J.Wu & F.H.Larsen, "Phase retrieval in near-field measurements by array synthesis", IEEE International Antennas Symposium Digest, pp.1474-1477, 1991.

[32] J.Skaat & H.E.Engan, "Phase reconstruction from reflectivity in fiber Bragg gratings", Optics Letters, Vol.24, No.3, pp.136-138, Feb.1999.

[33] A.Burian, P.Kuosmanen & C.Rusu, "1-D direct phase retrieval", Signal Processing, Vol.82, pp.1059-1066, 2002.

[34] A.Burian, P.Kuosmanen & C.Rusu, "1-D non-minimum phase retrieval by gain differences", pp.1001-1004, 1999.

[35] W.Kim & M.H.Hayes, "Iterative phase retrieval using two Fourier transform intensities", Proc.*ICASSPO*, pp.1563-1566, 1990.

[36] W.Kim & M.H.Hayes, "Phase retrieval using two Fourier transform intensities", J.Opt. Soc. Am., Vol.7, No.3, pp.441-449, March 1990.

[37] M.J.Bastiaans & K.B.Wolf, "Phase reconstruction from intensity measurements in linear systems", J.Opt. Soc. Am. A, Vol.20, No.6, pp.1046-1049, June 2003.

[38] O.M.Bucci, G.D'Elia, G.leone & R.Pierri, "Far-field pattern determination from the near-field amplitude on two surfaces", IEEE Trans. Antennas Propagat., Vol.AP-38, pp.1772-1779, Nov.1990.

[39] T.Isernia, R.Pierri & G.Leone, "New technique for estimation of far-field from near-zone phaseless data", Electronics Letters, Vol.27, pp.652-654, 1991.

[40] A.Tennant, G.Junkin & A.P.Anderson, "Antenna far-field predictions from two phaseless cylindrical near-field measurements", Electronics Letters, Vol.28, pp.2120-2122, 1992.

[41] O.M.Bucci, G.D'Elia & M.D.Migliore, "An effective near-field far-field transformation technique from truncated and inaccurate amplitude-only data", IEEE Trans. Antennas Propagat., Vol.47, pp.377-385, 1999.

[42] F.Las-Heras & T.K.Sarkar, "A direct optimisation approach for source reconstruction and NF-FF transformation using amplitude-only data", IEEE Trans. Antennas Propagat., Vol.50, pp.500-509, 2002.

[43] P.Hallbjörner, "Retrieving field amplitude and polarisation ellipse from phaseless measurements of linear polarisation, including error analysis", IEE Proc. Microwaves Antennas Propagat., Vol.150, pp.28-33, 2003.

[44] P.K.Koivisto & J.C.Sten, "Genetic algorithm applied to determine the spherical wave expansion from amplitude-only far-field data", Microwave & Optical Tech. Letters, Vol.46, No.4, pp.402-406, Aug.2005.

[45] S.Costanzo, G, di Massa & M.D.Migliore, "A novel hybrid approach for far-field characterization from near-field amplitude-only measurements on arbitrary scanning surfaces", IEEE Trans. Antennas & Propagation, Vol.53, No.6, pp.1866-1874, June 2005

Chapter 4

Model-Based Phase Retrieval

4.1 BASIC PRINCIPLES

We stated in Chapter 3 that, in the present one-dimensional case, where $H(\omega)$ is a function of a single variable ω, it can be shown mathematically that given $|H(\omega)|$ there is no unique phase $\arg\{H(\omega)\}$ associated with $H(\omega)$. In other words, it is not possible to retrieve a unique phase if only $|H(\omega)|$ has been specified. It is *always necessary* to specify additional constraints (i.e. more information about the transfer function) in order to retrieve $\arg\{H(\omega)\}$ from a given $|H(\omega)|$. The question remains as to what are the appropriate constraints in any particular case. The new method being proposed here for the retrieval of phase-related quantities from magnitude-only measurements on two-port microwave systems is based on the following argument : We assume that we are interested in the routine testing of a microwave network that is being manufactured in relatively large quantities. A reduced-order model is developed using the prototype of such a system. Such a reduced-order model will include a number of "coefficients" (other than frequency) whose value is determined using a full vector data set on the prototype. Establishing the form required for the reduce-order model acts as the additional constraints that are needed to provide a unique relationship between the transfer function magnitude and its phase. When later production models are tested using magnitude-only measurements, optimization is used (with the above-mentioned "coefficients" as the

optimization variables, for which only small adjustments are needed from one device to another) to match the measured and modelled magnitude responses. The retrieved "coefficients" are then used to predict the phase response using the model.

4.2 AVAILABLE MODELS OF ELECTROMAGNETIC SYSTEMS

All electromagnetic systems are properly modelled using electromagnetic theory, and electromagnetic simulations have become mainstream since commercial (and relatively reliable) software implementations of these methods are now widely available. However, designs are seldom done on the basis of electromagnetic simulations without having some form of simpler model in mind. By "simpler" we do not necessarily mean inferior or less accurate but rather something in which, over some restricted range of operation, only the response-determining features have been retained. These "simpler" (or reduced-order) models include equivalent circuit models [1,2,3,4], artificial neural network models [5,6,7] and rational function models [8-21]. In Section 4.3 we will apply model-base phase-retrieval using a rational function model. In Sections 4.4 and 4.5 we use equivalent circuit models in the model-based phase-retrieval process.

4.3 BRIDGED-T TWO-PORT CIRCUIT

4.3.1 Preliminary Remarks & Required Performance Parameters

We will establish the soundness of the proposed model-based phase retrieval algorithm using artificially generated "measured" data. The system considered will be the

bridged-T circuit shown in Figure 4.3-1. This has been selected purposely because it is known to be a non-minimum phase system. We can select a set of component values and use circuit analysis[8] to provide the "measured response" of the prototype. This can be used for set up the model for the system. Subsequent "production line" systems can be emulated by making sensible alterations to the element values and using circuit analysis to predict their "measured response"; only the magnitudes of the "measured" S-parameters of the "production-line" (i.e. non-prototype) systems will of course be used in the phase retrieval process. We emphasise that circuit analysis is done to generate the "measured" data only.

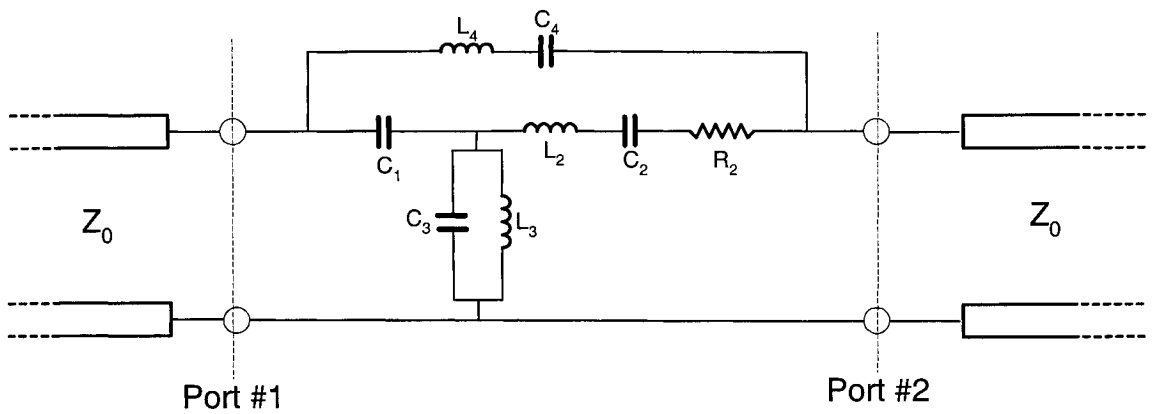


Figure 4.3-1 : Bridged-T Two-Port Circuit

4.3.2 Reduced-Order Modelling Considerations

The reduced-order model to be used in the model-based phase-retrieval algorithm will, in this section, be a rational function model obtained using the Cauchy model[16,17,19]. The transmission coefficient is expressed in the form

[8] The appropriate circuit analysis is detailed in Appendix II.

$$S_{21}(f) = \frac{\sum_{p=0}^{P} a_p f^p}{\sum_{q=0}^{Q} b_q f^q} \tag{4.3-1}$$

The method described in [19] can be used to determine the required P and Q, as well as to determine the coefficients $\{a_0, a_1, ..., a_P\}$ and $\{b_0, b_1, ..., b_Q\}$. We use the prototype system, whose measured S-parameters are known, to set up the Cauchy model for the system whose subsequent prototype models are to have their phase retrieved using their measured S-parameter magnitudes only. The above coefficients, determined from the prototype system, are used as starting values in the numerical optimization that forms part of the phase-retrieval algorithm.

4.3.3 Model-Based Phase Retrieval Algorithm

In order to describe the model-based phase-retrieval algorithm we must define an objective function. To do this we first define the partial objective function at the n-th frequency f_n as

$$F_{obj}^{(n)}\{a_0, a_1, ..., a_P, b_0, b_1, ..., b_Q\} =$$

$$20\log_{10}\left\{ \left| S_{21}^{Cauchy}\left(f_n, a_0, a_1, ..., a_P, b_0, b_1, ..., b_Q\right) \right| \right\} - 20\log_{10}\left\{ \left| S_{21}^{Meas}(f_n) \right| \right\} \tag{4.3-2}$$

and then the overall objective function is formed as

$$F_{obj}\{a_0, a_1, ..., a_P, b_0, b_1, ..., b_Q\} = \sum_{n=1}^{N} F_{obj}^{(n)}\{a_0, a_1, ..., a_P, b_0, b_1, ..., b_Q\} \tag{4.3-3}$$

Quantity $|S_{21}^{measure}(f)|$ is the measured S-parameter magnitude at a set of frequencies $\{f_1, f_2, ..., f_{N_f}\}$, of the system where phase is to be retrieved. The quantity $|S_{21}^{Cauchy}(f)|$ is the S-parameter magnitude that can be predicted using the Cauchy model if the coefficients $\{a_0, a_1, ..., a_P\}$ and $\{b_0, b_1, ..., b_Q\}$ are known. The model-based phase-retrieval procedure consists in finding the set of a_p and b_q values that minimize $F_{obj}\{...\}$. The sets of a_p and b_q values that minimize $F_{obj}\{...\}$ are then substituted in the Cauchy model to determine $\arg\{S_{21}(f)\}$, which is the retrieved phase. In the optimisation procedure the variables are $\mathrm{Re}\{a_p\}$ and $\mathrm{Im}\{a_p\}$ for $p = 0,1,.....,P$ and $\mathrm{Re}\{b_q\}$ and $\mathrm{Im}\{b_q\}$ for $q = 0,1,.....,Q$. We have used a direct search algorithm [22,23] for the minimisation of $F_{obj}\{a_0, a_1, ..., a_P, b_0, b_1, ..., b_Q\}$.The starting values for the a_p and b_q values are those determined from the system prototype, as remarked in Section 4.3.2. The range of the a_p and b_q values are determined from known fabrication tolerances in same production line process. Thus optimization algorithm will always provide the same solution for a given set of measured magnitude values.

4.3.4 Illustrative Example

We select the lower frequency as f_L = 0.1 GHz and the upper frequency f_U = 2.0 GHz. In order to illustrate one of a number of examples that were considered we select the component values shown in Table 4.3-1, and designate this to be the "prototype" network. In other words, we are able to use these actual components values and can use the circuit analysis of Appendix II to determine the S-parameters of the prototype in both magnitude and phase. This corresponds to being able to measure the S-parameters of a

network using a VNA, and we will call this the "complex measured data" for the prototype. This "complex measured data" for $\mathrm{Re}\{S_{21}^{Meas}(f)\}$ and $\mathrm{Im}\{S_{21}^{Meas}(f)\}$ is then used to establish the Cauchy model of the prototype; the coefficients of this model are provided in Table 4.3-2. The Cauchy model has P = 17 and Q = 18. A total of 41 frequency points were used in setting up the Cauchy model, with the frequency specified in GHz units. The success of the Cauchy modeling is illustrated in Figures 4.3-2 and 4.3-3.

Table 4.3-1 : Components Values of the Prototype Network

Component	Value
C_1	5pF
C_2	10pF
L_2	10nH
R_2	20Ω
C_3	5pF
L_3	10nH
C_4	10pF
L_4	15nH

Table 4.3-2 : Coefficients of the Cauchy Model of the Prototype Bridged-T Network

k	$a_k = (Re\{a_k\} , Im\{a_k\})$
0	(5.854326004039590E-010,0.000000000000000E+000)
1	(-1.426198908508523E-004,1.655569382146901E-004)
2	(-7.039107001988966E-003,-3.519195598308047E-003)
3	(1.718423956476306E-002,-2.141615168563588E-002)
4	(8.703844018958061E-002,6.617660682861928E-002)
5	(-0.144924676125515,0.154979378923237)
6	(1.885998374742002E-005,-0.175115988096449)
7	(-2.655783959512279E-002,-0.224519109372543)
8	(-7.434967963608355E-002,6.757370405492498E-002)
9	(5.665171381905518E-002,1.708240356352797E-002)
10	(1.765575871732020E-002,6.594115679872312E-004)
11	(7.323871582216594E-003,-1.166994535247707E-003)
12	(-1.119218989797071E-002,4.164590998848967E-003)
13	(4.229832757702229E-003,-9.638222052312221E-004)
14	(-6.450169988192713E-004,-2.390320244903506E-003)
15	(-1.150851641502750E-003,8.697722158280405E-004)
16	(6.219568037950507E-004,-1.248783182756136E-004)
17	(-8.502930760921349E-009,5.149765573867907E-008)

k	$b_k = (Re\{b_k\} , Im\{b_k\})$
0	(2.635038794005520E-005,2.269384337700776E-005)
1	(-8.096902727473765E-004,1.410150072295991E-003)
2	(-1.624684128175672E-002,-9.343298937870531E-003)
3	(5.091175614446537E-002,-7.256127786229657E-002)
4	(0.205019402908175,0.172200710672134)
5	(-0.358715958586670,0.342072116874941)
6	(-0.310995153748792,-0.447200840112691)
7	(0.255757194083311,-0.342609274814628)
8	(0.160135098531142,0.102361319885686)
9	(1.442365516502424E-002,3.853310448293729E-002)
10	(3.649351558824801E-002,2.500534256111098E-002)
11	(-1.295131883355405E-002,2.271547323331657E-002)
12	(-2.218173575332603E-002,8.828099218148569E-003)
13	(-3.042152060597381E-003,-1.196337271532963E-002)
14	(5.358024200558434E-003,-3.818070090553607E-004)
15	(-7.361266453947067E-004,8.074442093728606E-004)
16	(1.575568923176164E-004,-5.342931763446735E-004)
17	(6.305148001718788E-004,3.572432650470625E-004)
18	(-2.630590645484327E-004,5.231614016699110E-005)

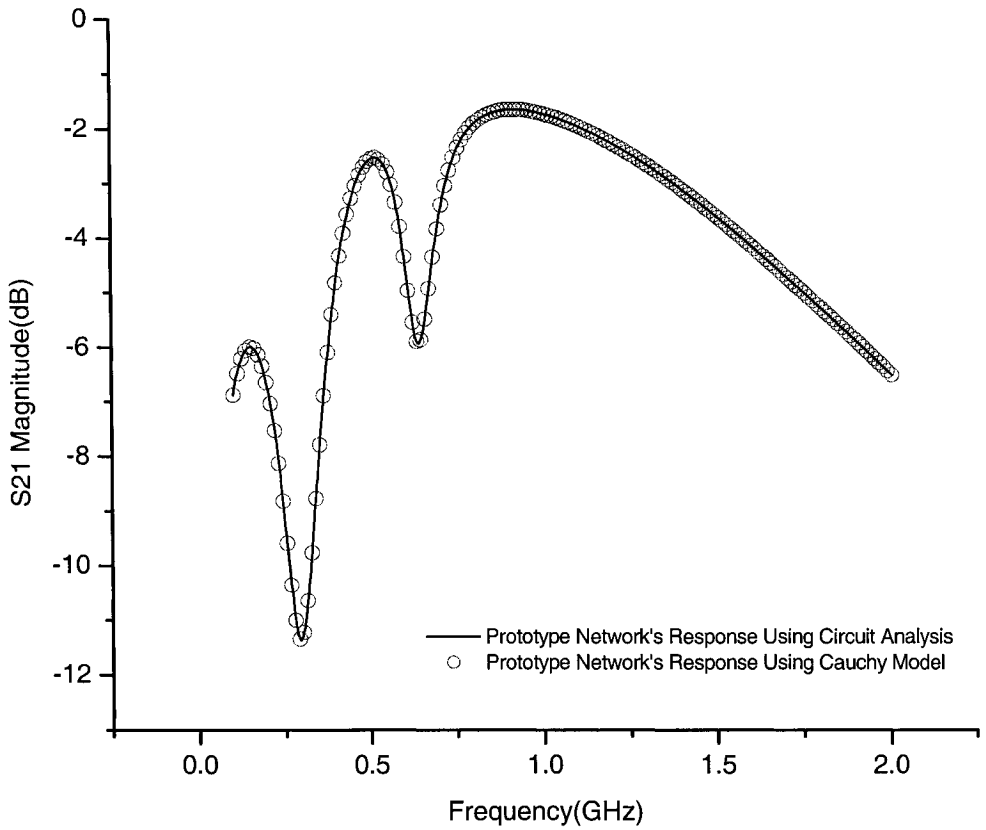


Figure 4.3-2 : "Measured" Magnitude $|S_{21}|$ of the prototype network and the $|S_{21}|$ predicted using the Cauchy model set up using the complex S_{21} of the prototype

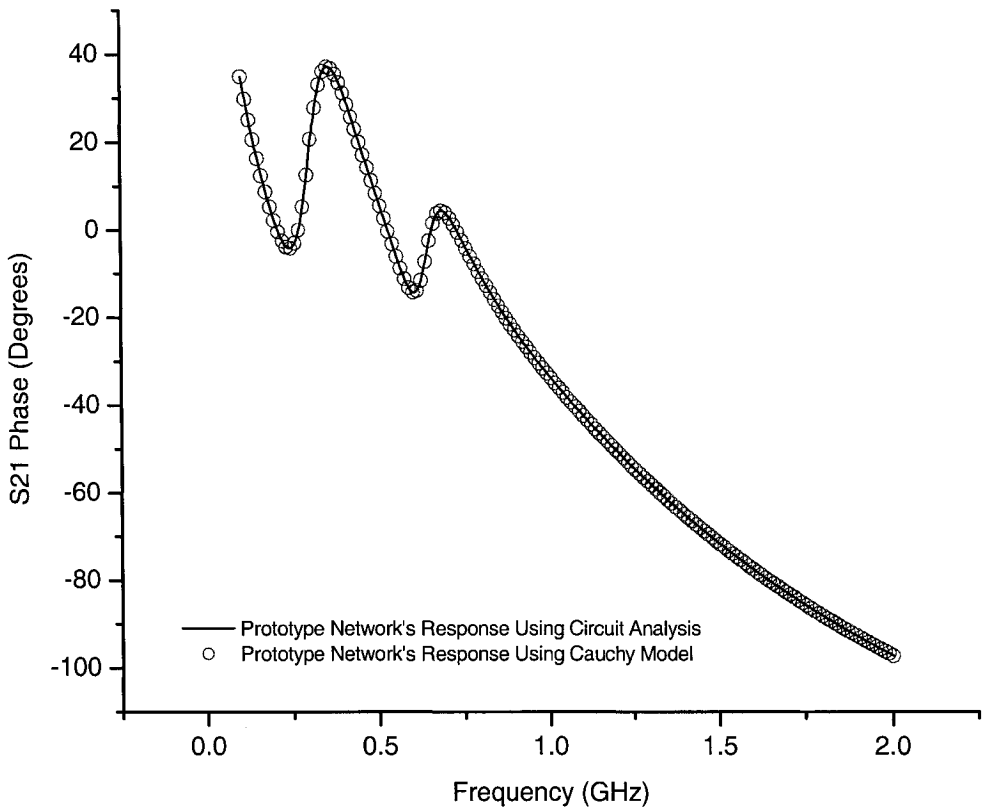


Figure 4.3-3 : "Measured" phase $\arg\{S_{21}\}$ of the prototype network and the $\arg\{S_{21}\}$ predicted using the Cauchy model set up using the complex S_{21} of the prototype

In order to emulate the production line process we alter the values of the components in Table 4.3-1. Bahl [24] indicates that lumped components typically have tolerances between ±5% and ±20%. We therefore increase C_4 by 20% to 6pF, increase L_3 by 20% to 12nH and decrease L_4 by 20% to 12nH. This provides us with a "production line" (i.e. non-prototype) sample of the network. We use circuit analysis with these altered component values to generate the "measured" values $|S_{21}^{measure}(f)|$ of the non-prototype network. Application of the phase-retrieval algorithm described in Section 4.3.3, using 801 frequency values, provides the retrieved phase shown in Figure 4.3-4. It is clear that the phase-retrieval process is very successful indeed. In order to provide some indication of algorithmic convergence we have plotted $F_{obj}\{...\}$ versus iteration number in Figure 4.3-6.

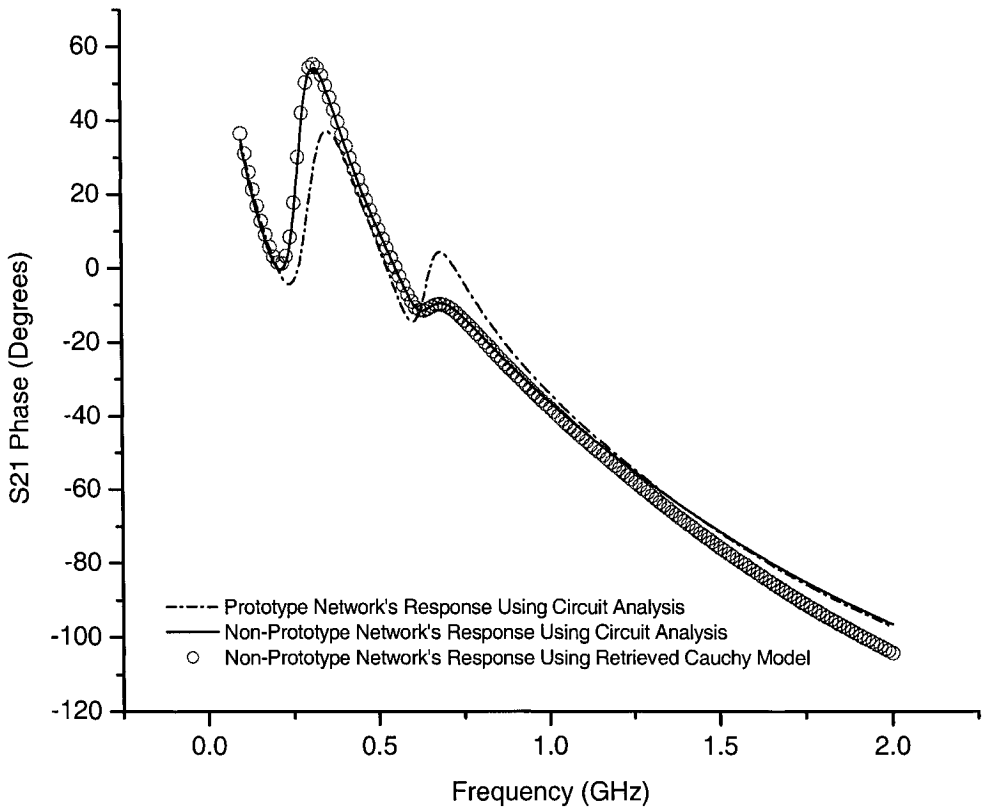


Figure 4.3-4: Retrieved $\arg\{S_{21}\}$ of the non-prototype network compared to its actual $\arg\{S_{21}\}$

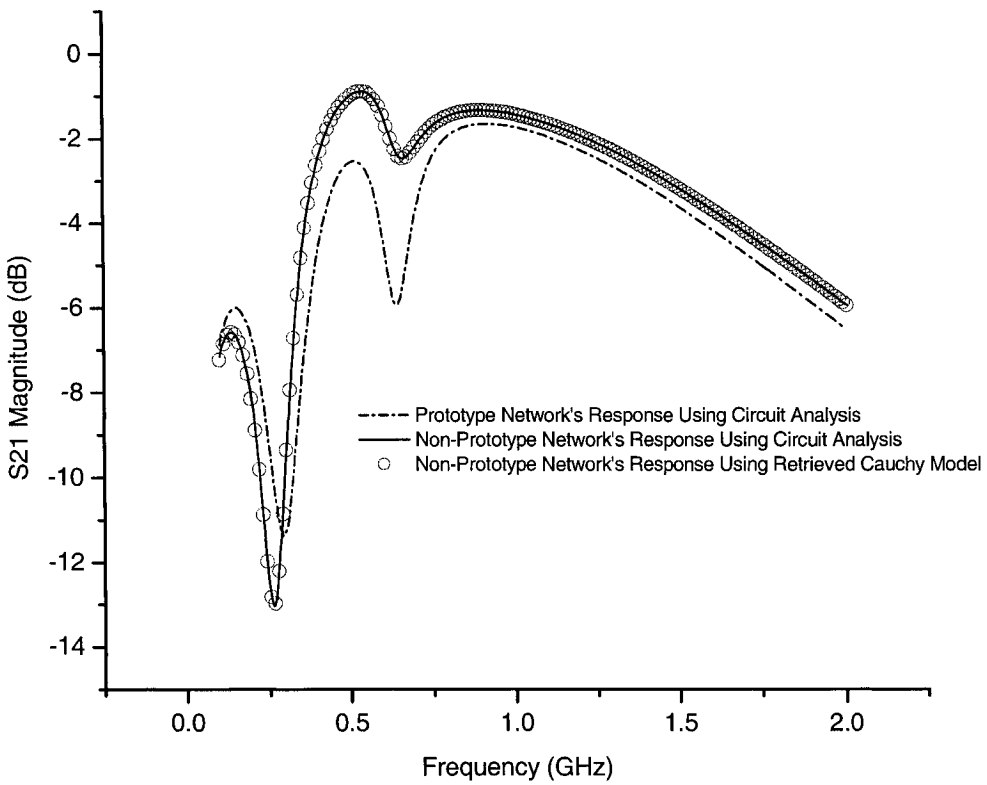


Figure 4.3-5 : Comparison of the "measured" magnitude $|S_{21}|$, and that of the retrieved model, of the non-prototype network

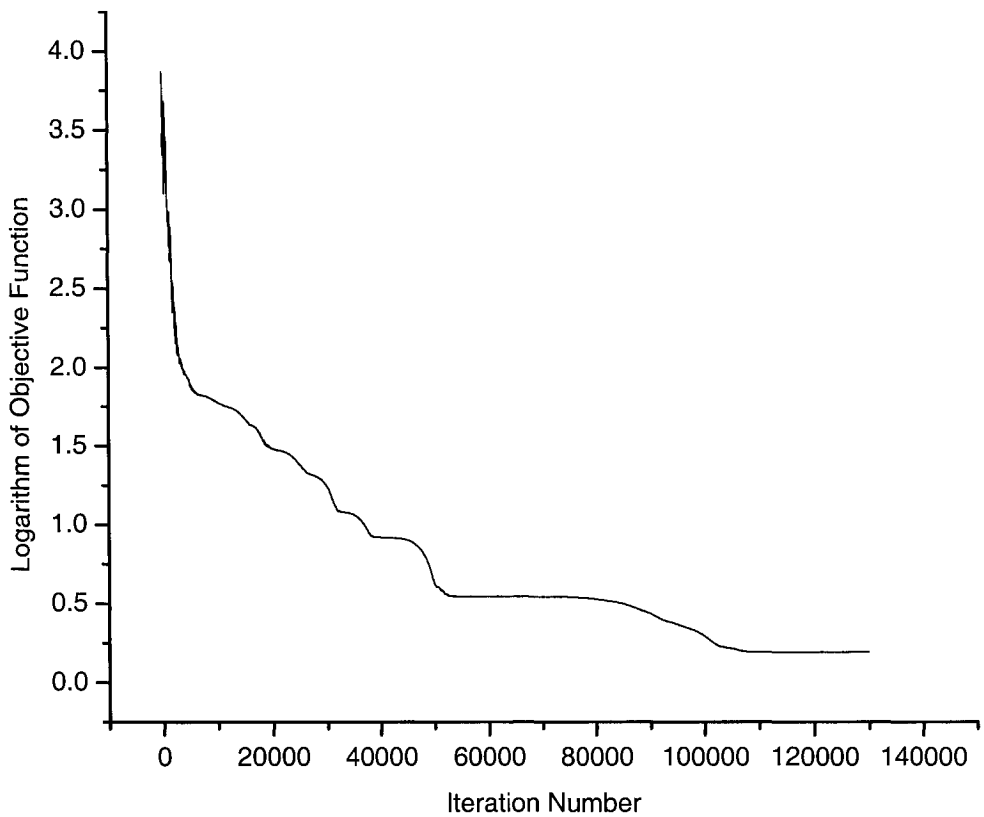


Figure 4.3-6 : The progress of $F_{obj}\{...\}$ values to the optimization algorithm iteration number

4.4 FIVE-POLE DIELECTRIC-RESONATOR-LOADED COUPLED CAVITY BANDPASS FILTER

4.4.1 Introductory Remarks

This example is the 5-pole dielectric resonator loaded coupled cavity bandpass filter described in [25] and shown in Figure 4.4-1. Due to the presence of the multiple paths from one port to the other this is clearly a non-minimum phase structure.

4.4.2 Equivalent Circuit Modelling Considerations

The reduced-order model most often used in the design of such filters is the equivalent circuit shown in Figure 4.4-2 [25,26], and this is therefore the sensible one to use in a model-based phase-retrieval algorithm. In Figure 4.4-2 the quantity N is the order of the filter, M_{ij} represents the coupling between resonators i and j, respectively, and R_i and R_j represent the input/output coupling. If the additional short lengths of waveguide at the input and output ends of the complete filter structure shown in Figure 4.4-1 were non-dispersive they would have little influence on the computed group delay[9]. In the present case the lengths of rectangular waveguide on either side of the filter are dispersive, and so will effect the group delay. We therefore include lengths of additional waveguide at the input and output of the filter (denoted ℓ_{IN} and ℓ_{OUT}) as shown in Figure 4.4-3.

The scattering parameter of the equivalent circuit model of Figure 4.4-2 are given by [27,28]

[9] In the case of the cable-wrap assembly discussed in Section 4.5 we will deal with a TEM waveguiding structure that has negligible dispersion. There, although the additional short lengths of transmission line due to the connector structure that exists beyond the discontinuities at each end add a small amount of additional phase shift to S_{21}, it has little influence on the group delay computed using $\phi(\omega) = \arg\{S_{21}\}$.

$$S_{21}' = -2j\sqrt{R_1 R_2}\{[\xi[I] - j[R] + [M]]^{-1}\}_{N,1} \quad (4.4\text{-}1)$$

$$S_{11}' = 1 + 2jR_1\{[\xi[I] - j[R] + [M]]^{-1}\}_{1,1} \quad (4.4\text{-}2)$$

In these expressions we have:

- $\xi = \frac{\omega_0}{\Delta\omega}(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega})$ where ω_0 is the centre frequency and $\Delta\omega$ is the bandwidth.
- $[I]$ is an $N \times N$ identity matrix
- $[R]$ is an $N \times N$ matrix of zero entries except for elements (1,1) and (N,N) which take the values R_1 and R_2 respectively.
- $[M]$ is the $N \times N$ coupling matrix of elements M_{ij}
- The notation $\{[A]\}_{mn}$ means "select the mn-th element of matrix [A]".

Using the known results for the changes in S-parameters due to a shift in reference planes [29], we have, for the complete equivalent circuit of Figure 4.4-3,

$$S_{21}^{model} = S_{21}' e^{-j\beta_{wg}(l_{in} + l_{out})} \quad (4.4\text{-}3)$$

$$S_{11}^{model} = S_{11}' e^{-j2\beta_{wg} l_{in}} \quad (4.4\text{-}4)$$

where $\beta_{wg} = \sqrt{\omega^2 \mu_0 \varepsilon_0 - (\frac{\pi}{a})^2}$, with a the larger dimension of the rectangular waveguides at the input/output ports of the filter in Figure 4.4-1.

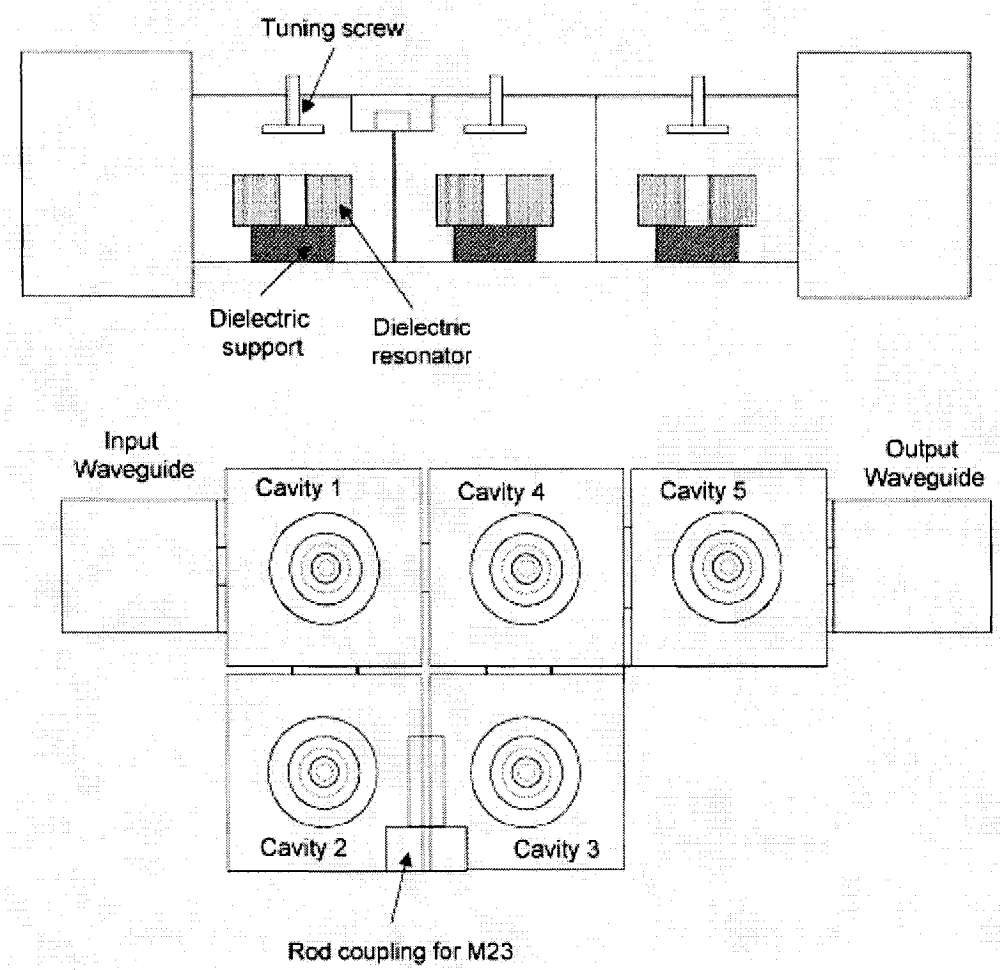


Figure 4.4-1 : Five-Pole Dielectric Resonator Filter Structure (After [25])

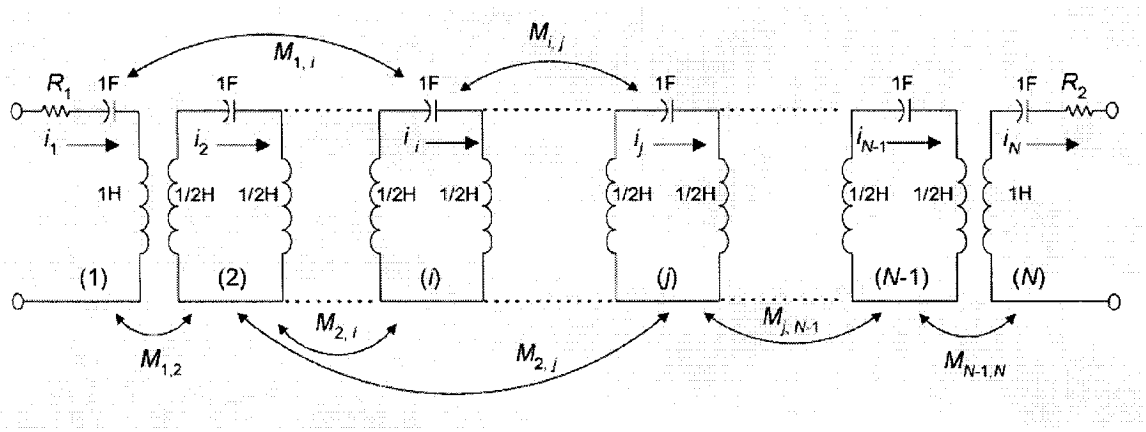


Figure 4.4-2 : Equivalent Circuit of an N-Coupled Resonator Filter (After [25])

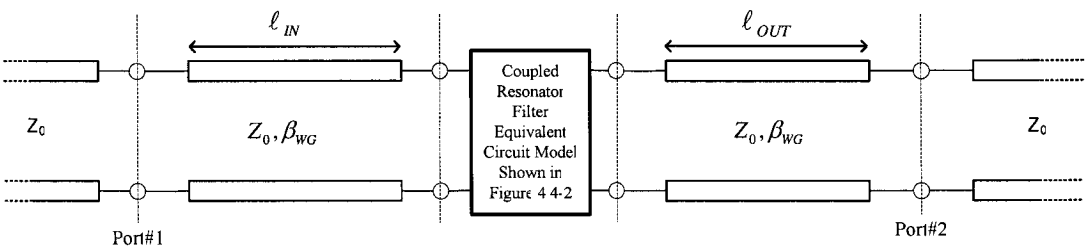


Figure 4.4-3 : Equivalent Circuit Model Used for Computation of the S-Parameters for Model-Based Phase-Retrieval

4.4.3 Model-Based Phase Retrieval Algorithm

The variables in the phase-retrieval algorithm are:

- The M_{ij}, j = 1,2...,N and i =1,2...,N
- R_1 and R_2

The objective function is

$$F_{obj}\{M_{ij}, i,j=1,2,...N; R_1, R_2\} = \sum_{n=1}^{N} F_{obj}^{(n)}\{M_{ij}, i,j=1,2,...N; R_1, R_2\} \quad (4.4\text{-}5)$$

where the partial objective function is

$$F_{obj}^{(n)}\{...\} = (20\log\{|S_{21}^{model}(...)|\} - 20\log\{|S_{21}^{measure}(...)|\})^2 + (20\log\{|S_{11}^{model}(...)|\} - 20\log\{|S_{11}^{measure}(...)|\})^2 \quad (4.4\text{-}6)$$

Using measured magnitude-only data $|S_{21}^{measure}(f_n)|$ and $|S_{11}^{measure}(f_n)|$ for n=1,2,..., N_f, a numerical optimization algorithm is used to minimize $F_{obj}\{...\}$. The values of R_1, R_2, l_{in}, l_{out} and $\{M_{ij}, i,j=1,2,...N\}$ which minimize $F_{obj}\{...\}$ are the retrieved values of these quantities for the particular filter that has been measured. Using these retrieved values, we then use the model (4.4-1) through (4.4-4) to retrieve $\arg\{S_{21}\}$ and $\arg\{S_{11}\}$.

4.4.4 Illustrative Example

In the case of a five-resonator filter, the ideal coupling terms (ideal in the sense that they will provide an ideal response) are:

$$[M]=\begin{bmatrix} 0.0 & 0.866 & 0.0 & -0.252 & 0.0 \\ 0.866 & 0.0 & 0.792 & 0.0 & 0.0 \\ 0.0 & 0.792 & 0.0 & 0.595 & 0.0 \\ -0.252 & 0.0 & 0.595 & 0.0 & 0.901 \\ 0.0 & 0.0 & 0.0 & 0.901 & 0.0 \end{bmatrix} \tag{4.4-7}$$

$$\begin{aligned} R_1 &= 1.133 \\ R_2 &= 1.133 \end{aligned} \tag{4.4-8}$$

This should be used [26] in the design of the coupling apertures between the resonators when designing an actual filter. Once the filter has been fabricated it will of course not exhibit the ideal response. We have obtained a complete set of measured S-parameters [30] that were also shown [25]. If the model-based retrieval algorithm, using magnitude-only data from the measured S-parameters, is now executed, we retrieve the following coupling parameter and input/output lengths :

$$[M]=\begin{bmatrix} 0.035678 & 0.878672 & -0.07258 & -0.24964 & 0.051768 \\ 0.878672 & 0.091155 & 0.793266 & -0.07758 & -0.00996 \\ -0.07258 & 0.793266 & 0.082853 & 0.612528 & -0.00312 \\ -0.24964 & -0.07758 & 0.612528 & 0.053485 & 0.948333 \\ 0.051768 & -0.00996 & -0.00312 & 0.948333 & 0.026716 \end{bmatrix} \tag{4.4-9}$$

$$\begin{aligned} R_1 &= 1.147696 \\ R_2 &= 1.211469 \end{aligned} \tag{4.4-10}$$

Although the measured magnitude data is also shown (in Figure 4.4-4 and Figure 4.4-6) it is the retrieved phase in Figure 4.4-5 and Figure 4.4-7 that is of principal interest. Using solely measured magnitude-only data it has indeed been possible to accurately retrieve

the phase. In the phase retrieval algorithm execution the ideal values of the coupling parameters were used as starting values.[10] If $|S_{22}|$ and $|S_{12}|$ are included in $F_{obj}\{...\}$ as well, there is very little difference in the outcome of the ...? retrieved ?.

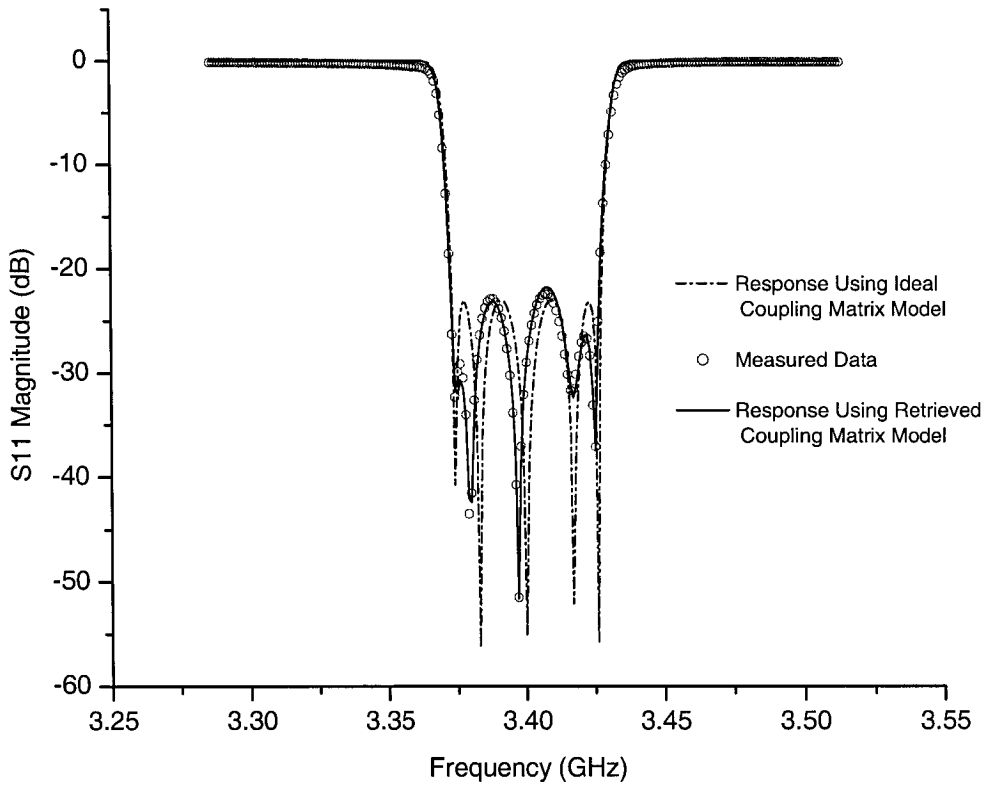


Figure 4.4-4 : Comparison of Measured and Retrieved $|S_{11}|$

[10] The starting values of l_{in} and l_{out} were not available. We therefore obtain these using the phase retrieval algorithm with complete measured complex S-parameter data by adjusting only l_{in} and l_{out}, but keeping the coupling parameters at their original values. In practice these values will be available from the design itself. The starting values are $l_{in} = 0.053909$ and $l_{out} = 0.035302$.

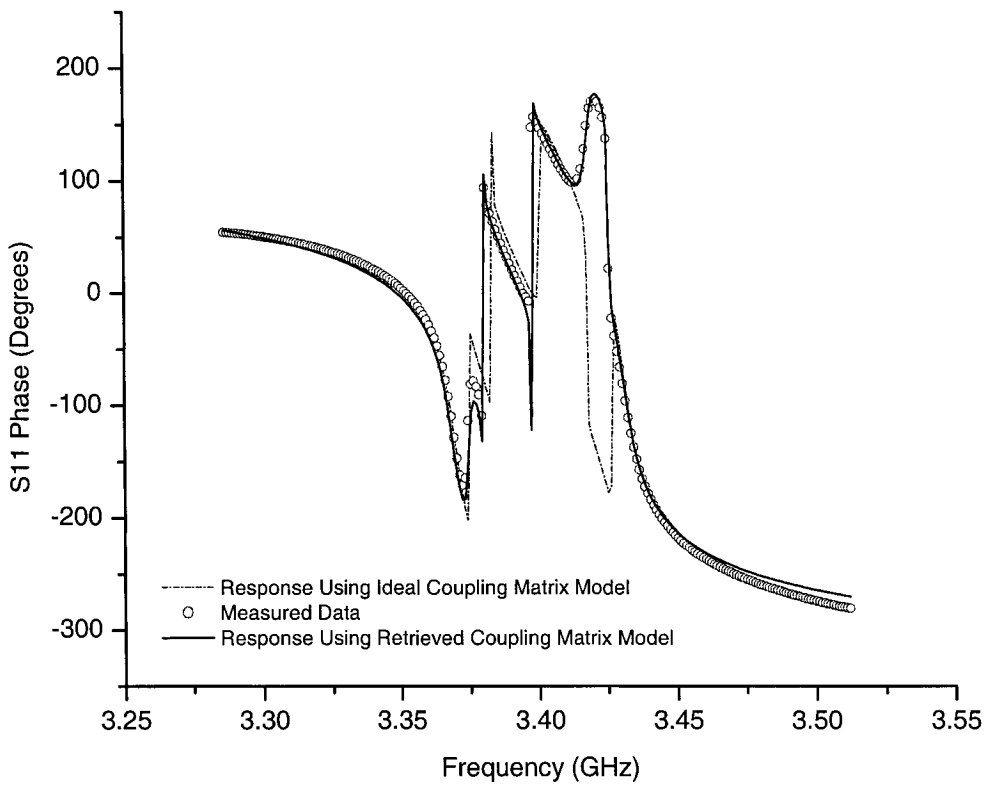


Figure 4.4-5 : Comparison of Measured and Retrieved $\arg\{S_{11}\}$

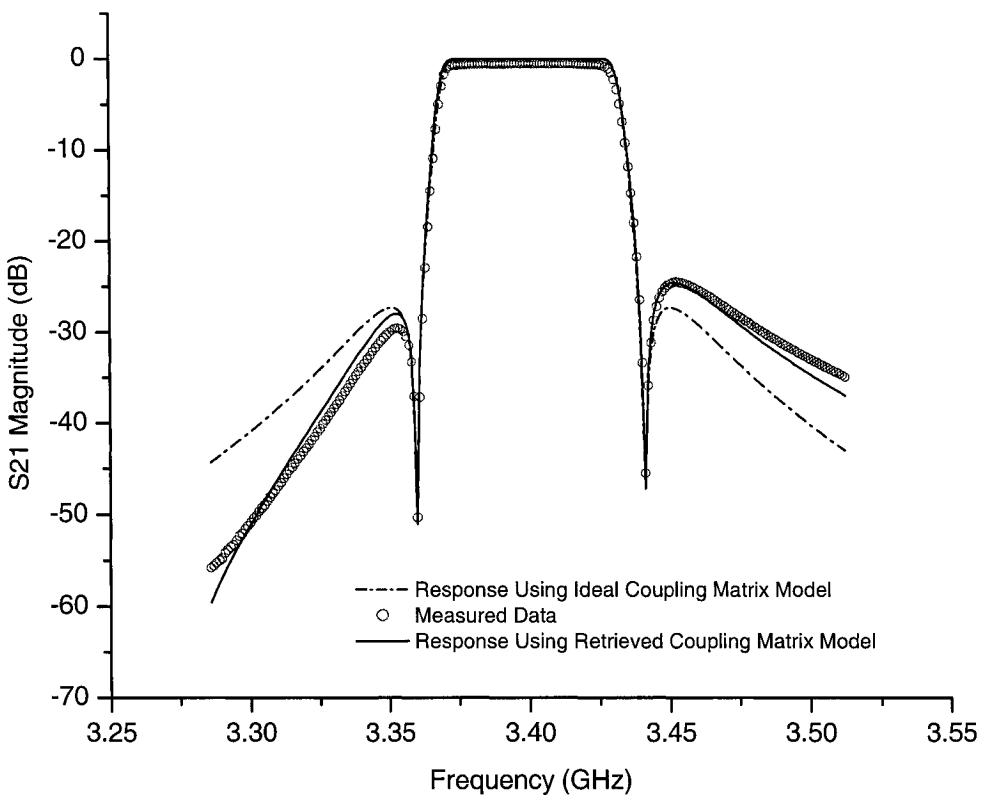


Figure 4.4-6 : Comparison of Measured and Retrieved $|S_{21}|$.

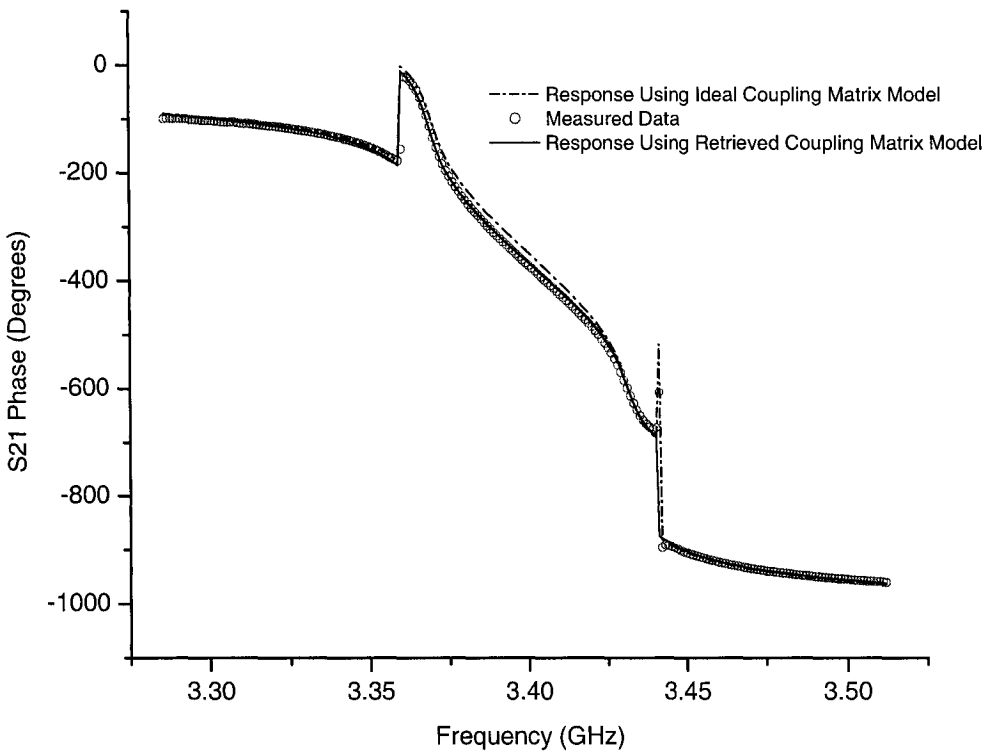


Figure 4.4-7 : Comparison of Measured and Retrieved $\arg\{S_{21}\}$

4.5 COAXIAL CABLE-WRAP ASSEMBLY FOR ANTENNA POINTING MECHANISMS

4.5.1 Preliminary Remarks & Required Performance Parameters

We have demonstrated the application of the model-based phase-retrieval approach in Sections 4.3 and 4.4. Now there is a class of problem for which the networks are electrically long and have a transmission coefficient $|S_{21}|$ always relatively close to unity but which exhibits a rapid but small "ripple" as a function of frequency. Yet it is important to be able to obtain $\arg\{S_{21}\}$ since in such cases it is the group delay variation and group delay slope that are also of interest. The "rapid but small" ripple contributes to the group delay variation. In particular, the ripple is not regular, in the sense that these are not all of equal amplitude. Any model used in the model-based phase-retrieval procedure must properly resolve this ripple.

One example of this class of problem is the cable-wrap assembly used in antenna pointing mechanisms. The cable-wrap assembly has a coaxial cable wrapped into the form of a coil (albeit with few turns). The coaxial cable has a connector at each end. The cable flexes as the antenna positioning mechanism experiences mechanical rotation, and hence its RF performance changes too. The RF specifications are usually such that each assembly needs to have its insertion loss and group delay variation (rather than the phase response itself) tested at each of a number of selected rotation points, and at different temperatures. Insertion loss can be found from a scalar measurement. However, we know from the performance definitions discussed in Section 2.3.4 that in order to know the group delay variation a vector measurement is needed, unless the scalar data can be used

to retrieve the group delay variation characteristics in some way. The goal of the present section is the development of a model-based algorithm for the retrieval of the group delay variation bounds (rather than the raw phase data) from scalar measurements. We do not know what random variation there will be in the cable braiding from one item to the next. The customisation will allow us to use a relatively simple equivalent circuit model that assumes an ideal cable and yet be able to retrieve the bounds on the group delay variation and slope.

Cable-wrap assemblies in antenna positioning mechanisms (APM) for space applications [31] undergo repeated RF testing for several reasons. The characteristics of a given cable can change according to the particular angular position of the APM, and due to a temperature change since such APMs typically operate of a temperature range -40°C to 80°C. RF testing is required at each stage of an ambient-cold-ambient-hot-ambient temperature cycle, which is possibly done more than once, with the APM in more than one angular position), and after one or more vibration tests. Thus RF testing is time-consuming. It would be good to minimise the time taken for such tests, and this requires that we simplify the test procedure as much as possible. As mentioned earlier direct phase measurement using a VNA is difficult due to cable flexing probhlems unless the VNA is dis-connected and re-connected each time the APM angular position needs to be changed. This is dreadfully time-consuming. And if it has to be done at hot or cold temperatures would be an unreasonable expectation. Magnitude-only scalar measurements would not only be more accurate if we had an appropriate group delay variation retrieval approach, but a significant time-saver.

As a matter of interest it is worth mentioning why such testing is usually considered necessary from one cable-wrap assembly to another. There will be slight changes in ε_{eff} and differences in the cable cross-sectional dimensions from one batch of cable to another. The connector dimensions will also be subject to such production variations. Furthermore, the assembly process of attaching the connectors to the flexible cables will result in small performance changes from one sample to another. Crimping can deform cables in an extremely variable manner. In spite of tight controls on the fabrication process there will be small differences in the length ℓ, and ℓ will alter as a result of the flexing which takes place due to rotation.

4.5.2 Equivalent Circuit Modelling Considerations

An equivalent circuit for the coaxial cable assembly is shown in Figure 4.5-1. We will attempt to motivate the choice of this equivalent circuit based on physical grounds [32,33].

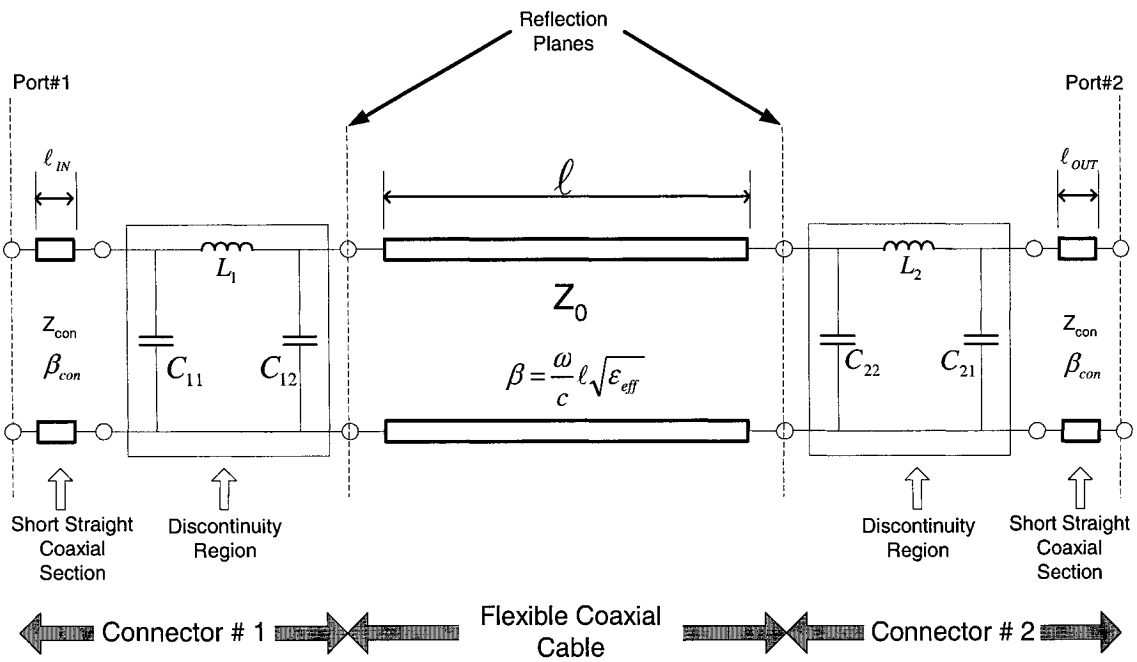


Figure 4.5-1 : Equivalent Circuit Model for a Coaxial Cable with Connectors

The connectors at the ends of the cable-wrap assembly are each modelled by a short section of transmission line of lengths ℓ_{IN} or ℓ_{OUT}, and a Π-network to model the physical discontinuities internal to the connector and at the connector/cable joint. The short section of transmission line has characteristic impedance Z_{con} and phase constant $\beta_{con} = (\omega/c)\sqrt{\varepsilon_c}$ where ε_c is the relative permittivity of the dielectric material in the connector. The C_{mn} are capacitances, where subscript *m* denotes the connector number (either 1 or 2) and subscript *n* identifies which capacitor we are talking about in the *m*-th connector. They do not denote mutual capacitances. Component L_m is an inductance. It is always possible to model a two-port as either a Π-network or a T-network. We have chosen a Π-network to model the portion of the cable-wrap assembly containing the discontinuities, and the elements of this network turn out, in the present case, to be capacitors and inductors as shown. The flexible coaxial cable itself is modelled as a section of transmission line of length ℓ, characteristic impedance Z_o, and phase constant $\beta = (\omega/c)\sqrt{\varepsilon_{eff}}$ where ε_{eff} is the effective relative permittivity of the dielectric material in the cable. We should not shirk from incorporating so-called "rules-of-thumb" in the model to be used. For instance, the experience of RF engineers is that losses are proportional to the square root of the frequency. We will therefore write the loss factor in the flexible coaxial cable as [37]

$$\alpha = \alpha_0 + \alpha_s\sqrt{f}$$

The term α_0 will determine the overall level of the loss and α_s its rate of change with frequency.

Computation of the S-parameters from the equivalent circuit model is performed using ABCD matrix cascading and converting the overall ABCD parameters to S-parameters. This is described in detail in Appendix I. We can denote the predicted responses by

$$S_{21}^{Model}\left\{\omega, C_{11}, C_{12}, L_1, C_{21}, C_{22}, L_2, \alpha_o, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}$$

and

$$S_{11}^{Model}\left\{\omega, C_{11}, C_{12}, L_1, C_{21}, C_{22}, L_2, \alpha_o, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}$$

Recall that we can predict both magnitude and phase, but assume throughout that we are only able to measure the magnitudes $\left|S_{11}^{Meas}(\omega)\right|$ and $\left|S_{21}^{Meas}(\omega)\right|$.

4.5.3 Customized Model-Based Phase Retrieval Algorithm

The coaxial cable is a true TEM structure, and so we would not expect any significant variation of ε_{eff} with frequency. However, as the cable flexes during rotation the value of $\ell\sqrt{\varepsilon_{eff}}$ can change. Furthermore, the cable-wrap assembly has to be tested not only at ambient temperature but also at cold and hot. The length ℓ and the relative permittivity ε_{eff} can change with temperature. Thus we need to be able to include $\ell\sqrt{\varepsilon_{eff}}$ as an unknown to be estimated for each new measurement. The detailed dimensions of the coaxial connector geometries can also change over temperature. Thus it is clear that all the equivalent circuit parameters have to be retrieved for each new measurement.

The implementation of the complete algorithm will be described in four steps, each in one of the four sub-sections that follow. The detail in which these will be described might create the impression that the algorithm is a computationally cumbersome one. However, the reader will observe that the actual arithmetic operations are very simple ones, and it is relatively easy to code the process. The success of the algorithm rests on its physical basis. With each step we will display numerical results using a particular cable-wrap assembly measured at room temperature ("ambient" temperature), which we will refer to as the "example cable-wrap assembly". Results for other assemblies will be presented in Section 4.5.8, after the retrieval algorithm has been thouroughly explained. We will also assume that the connectors are identical. In other words we here have $C_{11} = C_{21} = C_1$, $C_{12} = C_{22} = C_2$ and $L_1 = L_2$ and will write the modelled responses as

$S_{21}^{Model}\left\{\omega, C_1, C_2, L_1, \alpha_o, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}$ and $S_{11}^{Model}\left\{\omega, C_1, C_2, L_1, \alpha_o, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}$.

A note on some terminology will be helpful for clarity in later sections. In each of the algorithm's steps we will determine the final value of certain equivalent circuit parameters, which we will refer to as the retrieved values. In succeeding steps these parameters are always used in the model with their values set equal to their retrieved values. However, in steps preceding that in which a parameter is retrieved, we will find it useful to use the parameter in a way that does not retrieve its value but allows that of another parameter to be retrieved. This will become clearer when the algorithm is discussed in detail in Sections 4.5.4 through 4.5.7.

One fortunate aspect is that we will usually know the range of variation of many of the components in the equivalent circuit. Bounds can therefore be set on these quantities whenever they are used in optimisation algorithms; this is another manifestation of the

principle of using all the information available for the retrieval of phase-related data. Cable manufacturers will usually know (from their own fabrication process control knowledge) what the variation in ε_{eff} will be. By knowing what the largest acceptable return loss will be it is possible to estimate the largest values of C_1, C_2 and L_1.

The question might be asked as to what is to be done of a particular device under test has been so badly assembled that its equivalent circuit parameters lie outside the ranges set as bounds in a retrieval algorithm. Fortunately, in such cases the magnitude data itself (eg. insertion loss and return loss) will likely be such that the device is out of specification, and so phase-related retrieval will not be needed anyway.

4.5.4 Retrieval of Equivalent Circuit Parameter $\ell\sqrt{\varepsilon_{eff}}$

By analogy with Fabry-Perot resonators with partially-reflecting planes [34] there will be an underlying periodicity in the RF performance that is dependent on the electrical length $(\omega/c)\ell\sqrt{\varepsilon_{eff}}$ of the coaxial cable between the "reflection planes" shown. While the levels of the actual maxima and minima of the periodic response will depend on the values of the equivalent circuit elements C_1, C_2 and L_1 used to model the effects of the connectors, the frequency values at which these occur will not. We can thus retrieve the value of $\ell\sqrt{\varepsilon_{eff}}$ by examining this periodicity. The representative behaviour of S_{11} and S_{21} in Fig.4.5-2 and Fig.4.5-3 at once reveals that $\left|S_{11}^{Meas}\right|$ shows this periodicity in a more marked manner than $\left|S_{21}^{Meas}\right|$. We will therefore use the measured $\left|S_{11}^{Meas}\right|$ to determine $\ell\sqrt{\varepsilon_{eff}}$.

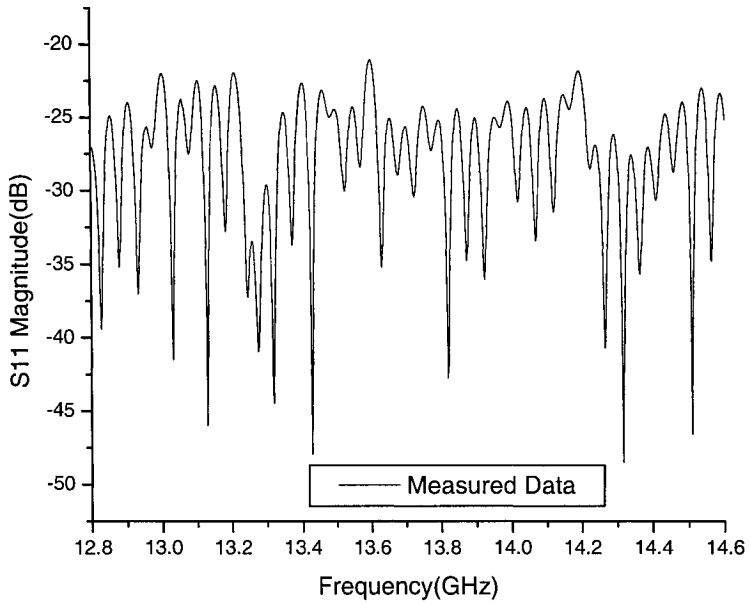


Figure 4.5-2 : Representative Measured Scattering Parameter Magnitude $|S_{11}^{Meas}|$ for an Actual Cable-Wrap Assembly

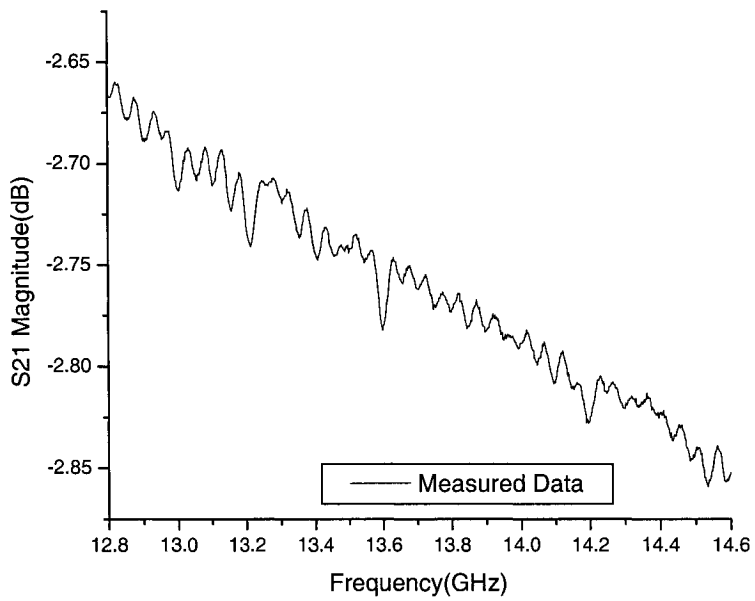


Figure 4.5-3 : Representative Measured Scattering Parameter Magnitude $|S_{21}^{Meas}|$ for an Actual Cable-Wrap Assembly

Note that the vertical scale for $|S_{21}^{Meas}|$ is greatly expanded.

Observe that while the sections of transmission line (of length ℓ_{IN} and ℓ_{OUT}) in the connectors beyond the Π-networks will influence the phase $\arg\{S_{21}\}$, the ripple is due to the length of cable between the reflection planes only. We therefore, in the entire retrieval algorithm, use the "reduced" equivalent circuit shown in Fig.4.5-4.

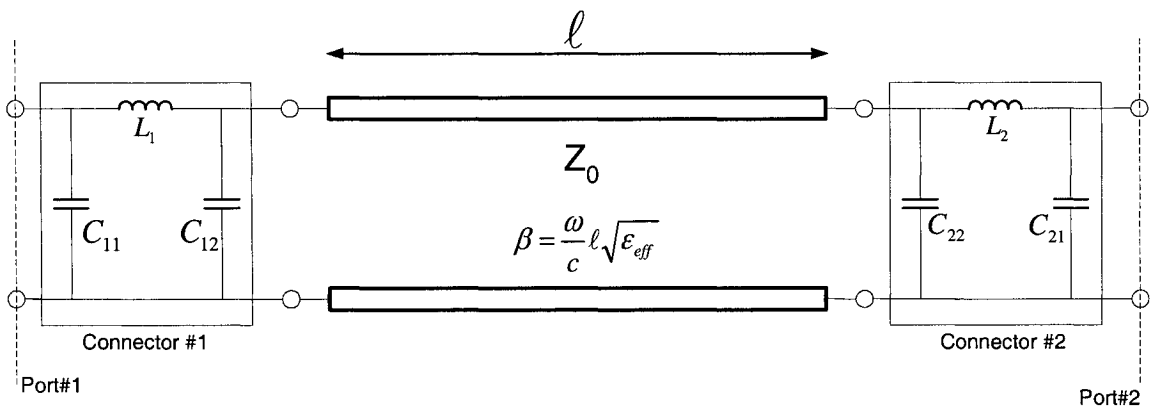


Figure 4.5-4 : Reduced Equivalent Circuit Model for a Coaxial Cable with Connectors

In this step we need to have the equivalent circuit model exhibit the periodicity of the response (i.e. it must reveal the "ripple"), and so we must simply select non-zero values of C_1, C_2 and L_1. The actual values selected are not important. We can also set $\alpha_o = 0$ and $\alpha_s = 0$. Since all equivalent circuit quantities other than $\ell\sqrt{\varepsilon_{eff}}$ are thus fixed during execution of this step of the retrieval algorithm we will write

$$\left|S_{11}^{Model}\left\{\omega, C_1, C_2, L_1, \alpha_o, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}\right| = \left|S_{11}^{Model}\left\{\omega, \ell\sqrt{\varepsilon_{eff}}\right\}\right|$$

We form the objective functions

$$F_{obj}^{(1)}\{\omega\} = \text{AverageSpacing Between Peaks of}\left\{\left|S_{11}^{Model}\left(\omega, \ell\sqrt{\varepsilon_{eff}}\right)\right|\right\} - \text{AverageSpacing Between Peaks of}\left\{\left|S_{11}^{Meas}(\omega)\right|\right\}$$

$$F_{obj}^{(2)}\{\omega\} = \text{No. of Peaks of}\left\{\left|S_{11}^{Model}\left(\omega, \ell\sqrt{\varepsilon_{eff}}\right)\right|\right\} - \text{No. of Peaks of}\left\{\left|S_{11}^{Meas}(\omega)\right|\right\}$$

$$F_{obj}^{(3)}\{\omega\} = \sum_{m=1}^{N}\left|\omega_{\text{m-th Peak}}^{\text{Meas}} - \omega_{\text{m-th Peak}}^{\text{Model}}\right|$$

Note that $\omega_{m-th\,Peak}$ is the frequency at which the m-th peak occurs. By "peak" we mean maxima or "dips" (i.e. minima) in the response. We adjust $\ell\sqrt{\varepsilon_{eff}}$ so that $F_{obj}^{(1)}\{\omega\} = 0$, $F_{obj}^{(2)}\{\omega\} = 0$ and $F_{obj}^{(3)}\{\omega\}$ is minimised. Fig.4.5-5 compares $\left|S_{11}^{Model}\right|$ to $\left|S_{11}^{Meas}\right|$. The procedure has aligned the maxima and minima of the measured and modelled quantities, and the value of $\ell\sqrt{\varepsilon_{eff}}$ which achieves this is the retrieved value of $\ell\sqrt{\varepsilon_{eff}}$. At this point we have used assumed values of the other equivalent circuit quantities, and so the levels are not yet correct. This will be taken care of as we determine the retrieved values of the remaining equivalent circuit quantities in the algorithmic steps that follow.

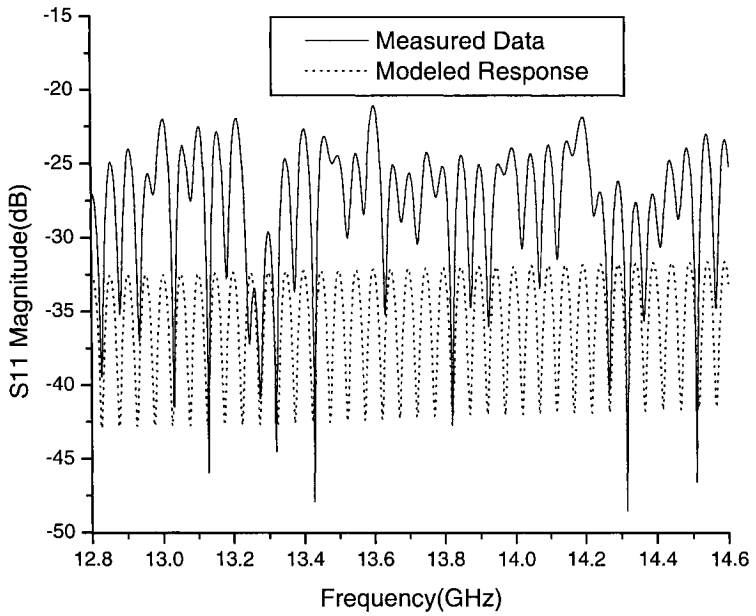


Figure 4.5-5 : Comparison of $|S_{11}^{Meas}|$ and $|S_{11}^{Model}|$ Using the Retrieved $\ell\sqrt{\varepsilon_{eff}}$ But Assumed Values of the Other Equivalent Circuit Quantities

It will be worthwhile to make a brief digression at this point to provide some re-assuring confirmation of the equivalent circuit argued in Section 4.5.2. The full complex S_{21}^{Meas} for the example cable-wrap assembly, obtained using a VNA, is plotted as $\mathrm{Re}\{S_{21}^{Meas}\}$ and $\mathrm{Im}\{S_{21}^{Meas}\}$ in Fig.4.5-6. Due to the fact that the cable–wrap has relatively small values of $|S_{11}^{Meas}|$ and $|S_{22}^{Meas}|$ a very approximate model for S_{21} (too approximate to allow us to determine the group delay variations, but sufficient for this digression) is simply

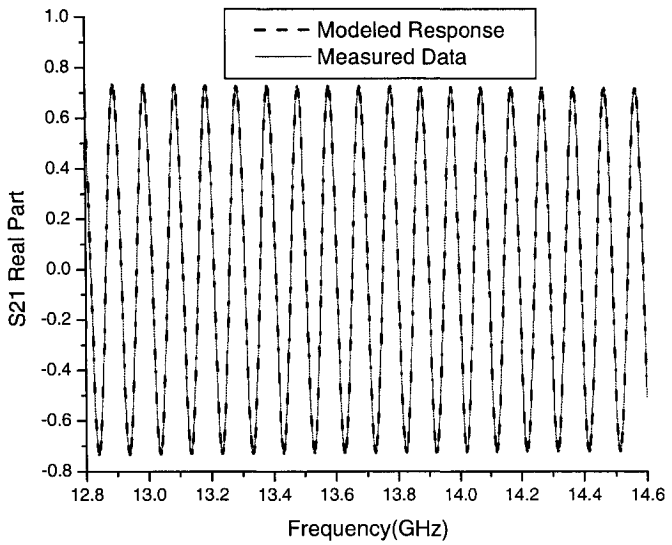
$$S_{21} = e^{-j\frac{\omega}{c}(\ell_{IN}+\ell+\ell_{OUT})\sqrt{\varepsilon_{eff}}} = e^{-j\frac{\omega}{c}\ell_{TOTAL}\sqrt{\varepsilon_{eff}}}$$

where we have assumed $\varepsilon_{con} = \varepsilon_{eff}$. This means that

$$\mathrm{Re}\{S_{21}\} = \cos\left(\frac{\omega}{c}\ell_{TOTAL}\sqrt{\varepsilon_{eff}}\right)$$

$$\mathrm{Im}\{S_{21}\} = -\sin\left(\frac{\omega}{c}\ell_{TOTAL}\sqrt{\varepsilon_{eff}}\right)$$

These have also been plotted in Fig.4.5-6, but for an incorrect value of $\ell_{TOTAL}\sqrt{\varepsilon_{eff}}$. In the case of the cable-wrap assembly being used as the illustrative example, adjustment of $\ell_{TOTAL}\sqrt{\varepsilon_{eff}}$ so that the measured and (approximately) modelled curves in Fig.4.5-6 are aligned yields $\ell_{TOTAL}\sqrt{\varepsilon_{eff}}$ =3.18. If we compare this to the value $\ell_{TOTAL}\sqrt{\varepsilon_{eff}}$ obtained earlier in this section, use the nominal connector ε_{eff} =2.54, and assume that the connectors are identical with $\ell_{IN} = \ell_{OUT}$, this gives $\ell_{IN} = \ell_{OUT} \approx 7.4mm$. This is indeed close to the physical lengths at the ends of the connectors.



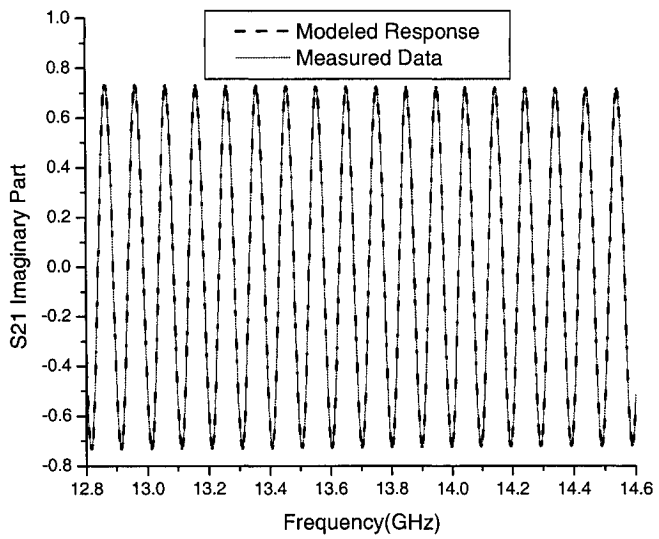


Figure 4.5-6 : Comparison of Measured and Modelled $\mathrm{Re}\{S_{21}\}$ and $\mathrm{Im}\{S_{21}\}$ Using the Retrieved Value of $\ell_{TOTAL}\sqrt{\varepsilon_{eff}}$ in the Model But Assumed Values for the Other Equivalent Circuit Parameters

4.5.5 Retrieval of Equivalent Circuit Parameter α_s

In this step we set $C_1 = C_2 = L_1 = \alpha_o = 0$, and keep the quantity $\ell\sqrt{\varepsilon_{eff}}$ fixed at its value determined in the preceding step. Only ω and α_s are variables, and so we will write

$$S_{21}^{Model}\left\{\omega, C_1 = 0, C_2 = 0, L_1 = 0, \alpha_o = 0, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\} = S_{21}^{Model}\left\{\omega, \alpha_s\right\}$$

Setting the components of the Π-networks to zero removes the effects of the connectors from the model. Furthermore, with $\alpha_o = 0$ the predicted $S_{21}^{Model}\left\{\omega, \alpha_s\right\}$ will not have the correct level. However, the goal in this step is to determine the quantity α_s, and its value is the dominant factor in determining the rate of decrease of $\left|S_{21}\right|$ with increasing frequency rather than the correct level. We wish to adjust α_s in such a way that, with the other equivalent circuit quantities eventually taking on their correct values, it will result in $\left|S_{21}^{Model}(\omega)\right|$ possessing the required downward trend with increasing frequency. Use of the equivalent circuit model (albeit with the connectors removed) ensures that we are at least establishing the value using the expected physical behaviour and not simply a straight line.

The procedure followed in this step is best described with reference to Fig.4.5-7. In addition to a typical $\left|S_{21}^{Meas}\right|$ there is $\left|S_{21}^{Model}\left\{\omega, \alpha_s = \alpha_s^{(k)}\right\}\right|$ labelled as Curve#1 and $\left|S_{21}^{Model}\left\{\omega, \alpha_s = \alpha_s^{(k+1)}\right\}\right|$ denoted as Curve#2.

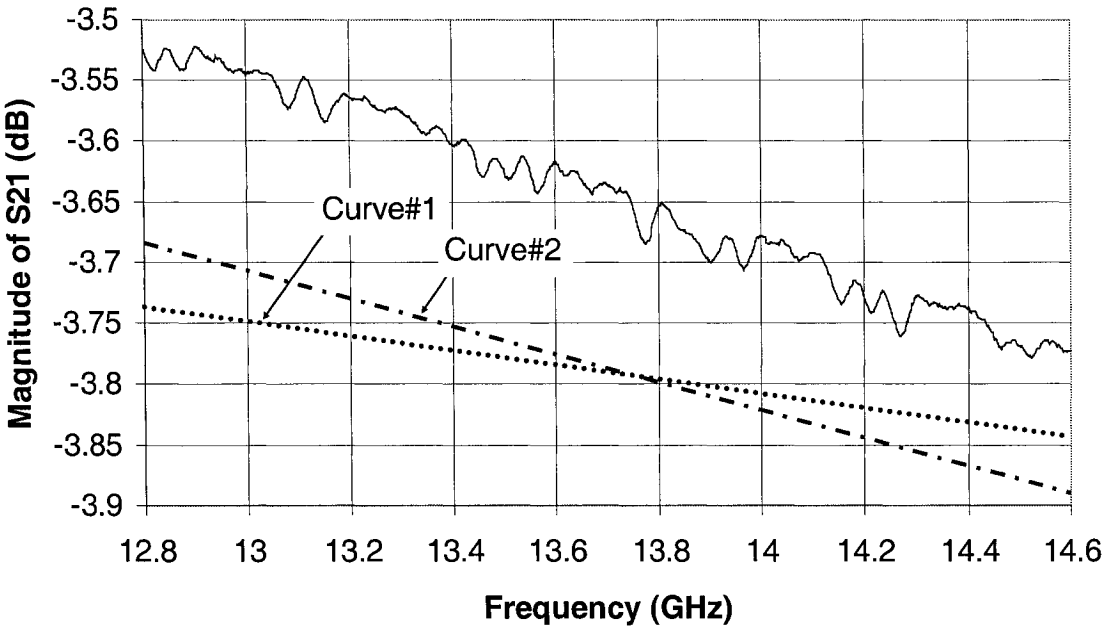


Figure 4.5-7 : Sketch of a Typical Response Mag{S_{21}}

The value of α_s is determined through minimisation of the objective function

$$F_{obj}\{\alpha_s\} = \sum_{n=1}^{N} \{P_n(\omega_n, \alpha_s) - P_{avg}(\alpha_s)\}$$

In this expression the term

$$P_n(\omega_n, \alpha_s) = \left| S_{21}^{Meas}\{\omega_n\} \right| - \left| S_{21}^{Model}\{\omega_n, \alpha_s\} \right|$$

and its average value over all frequencies is

$$P_{avg}(\alpha_s) = \frac{1}{N} \sum_{n=1}^{N} P_n(\omega_n, \alpha_s)$$

Using the above expressions we can write the objective function as

$$F_{obj}\{\alpha_s\} = \sum_{n=1}^{N}\left(P_n(\omega_n, \alpha_s) - \frac{1}{N}\sum_{m=1}^{N} P_m(\omega_m, \alpha_s) \right)$$

and its minimization supplies the required value of α_s, resulting in Fig.4.5-8.

Remember that only $\ell\sqrt{\varepsilon_{eff}}$ and α_s have been retrieved thus far, and so the curves are not yet aligned.

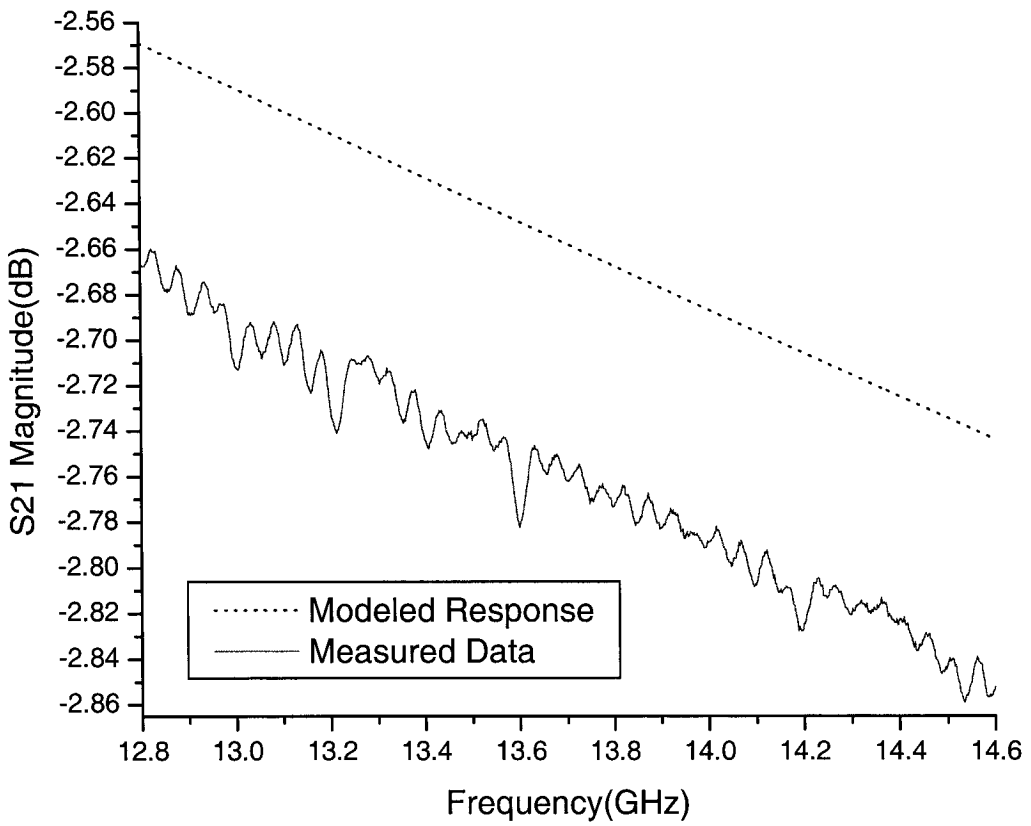


Figure 4.5-8 : Partially Retrieved $|S_{21}|$

4.5.6 Retrieval of Equivalent Circuit Parameters C_1, C_2 and L_1

We retrieve the values of C_1, C_2 and L_1 by adjusting these component values so that they result in $\left| S_{21}^{Model} \right|$ having the same upper and lower bounding curves as $\left|S_{21}^{Meas}\right|$. In other words, we simply wish these component values to provide the "correct ripple" and then in the final step in Section 4.5.8 we will adjust α_o to move $\left| S_{21}^{Model} \right|$ to the same overall level as $\left|S_{21}^{Meas}\right|$.

Using $\left|S_{11}^{Meas}\right|$ we earlier determined the frequency locations of the peaks, namely the $\omega_{m-th\,Peak}^{meas}$. We now read off the values of $\left|S_{21}^{meas}(\omega_{m-th\,Peak}^{meas})\right|$ and determine, for all values of m, the quantities

$$P_m^+ = \left| \left|S_{21}^{meas}(\omega_{m+1}^{meas})\right| - \left|S_{21}^{meas}(\omega_m^{meas})\right| \right|$$

and

$$P_m^- = \left| \left|S_{21}^{meas}(\omega_m^{meas})\right| - \left|S_{21}^{meas}(\omega_{m-1}^{meas})\right| \right|$$

The ripple amplitude in $\left|S_{21}^{Meas}\right|$ is then determined as

$$Ripple\left|S_{21}^{Meas}\right| = Max\left\{P_m^+, P_m^-, \forall m\right\}$$

We note that we are similarly able to write, using the model,

$$Ripple\left|S_{21}^{Model}\left\{\omega, C_1, C_2, L_1\right\}\right| = Ripple\left|S_{21}^{Model}\left\{\omega, C_1, C_2, L_1, \alpha_o = 0, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}\right|$$

We then adjust the values of C_1, C_2 and L_1 so that the measured and modelled ripples are the same. These are then the extracted values of C_1, C_2 and L_1, and provide Fig.4.5-9. Remember that only $\ell\sqrt{\varepsilon_{eff}}$, α_s, C_1, C_2 and L_1 have been retrieved thus far, and so the curves are not yet aligned.

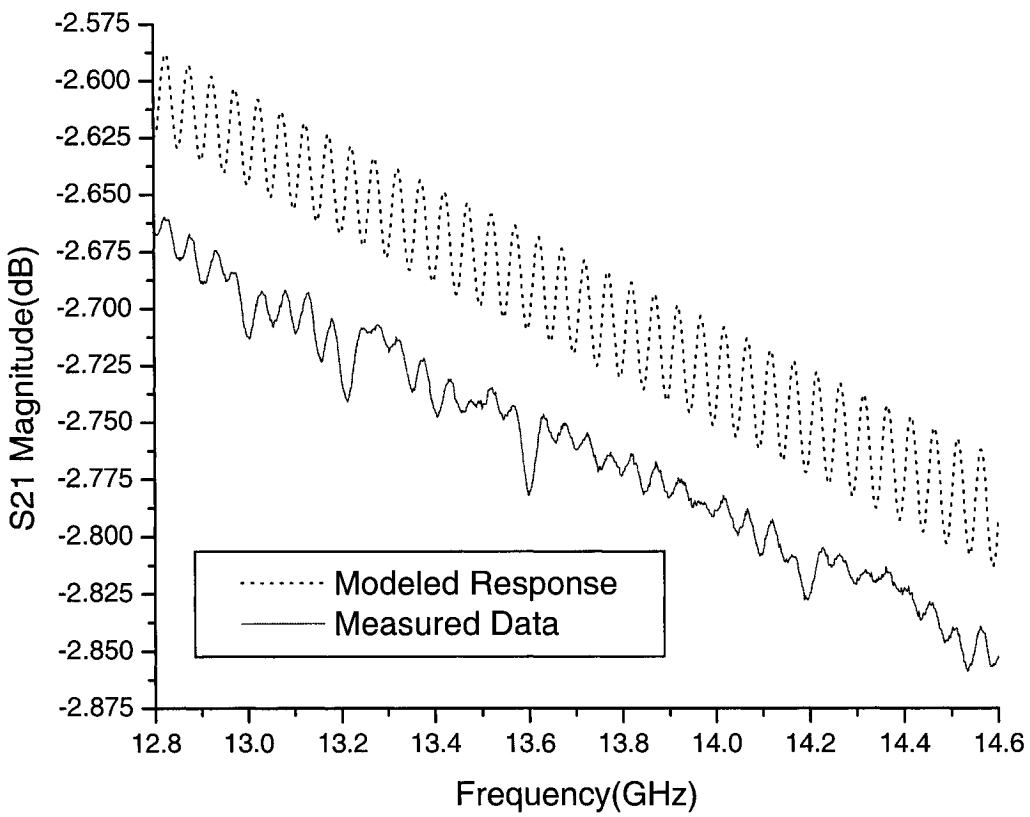


Figure 4.5-9 : Partially Retrieved $|S_{21}|$

4.5.7 Retrieval of Equivalent Circuit Parameter α_o

A. Determination of $f_{centre}(\omega)$

In this first part of this step the already-retrieved values of α_s and $\ell\sqrt{\varepsilon_{eff}}$ are used to form the function

$$f_{centre}(\omega, \alpha_{centre}) = \left| S_{21}^{Model}\left\{\omega, C_1 = 0, C_2 = 0, L_1 = 0, \alpha_o = \alpha_{centre}, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\}\right|$$

in which we consider α_{centre} to be a variable which must be determined. We form the objective function

$$F_{obj}\{\alpha_{centre}\} = \sum_{n=1}^{N} Q_n(\omega_n, \alpha_{centre})$$

where

$$Q_n(\omega_n, \alpha_{centre}) = \left|\, \left| S_{21}^{Meas}\{\omega_n\}\right| - f_{cenre}(\omega_n, \alpha_{centre}) \,\right|$$

Then we adjust α_{centre} until the objective function is minimised and fix it at this value. Thus $f_{centre}(\omega, \alpha_{centre})$ becomes a completely known function, which we can write as $f_{centre}(\omega)$, and is related to $\left|S_{21}^{Meas}\right|$ as illustrated in Fig.4.5-10.

B. Retrieval of α_o

Using $\left|S_{11}^{Meas}\right|$ we earlier determined the frequency locations of the peaks, namely the $\omega_{m-th\,Peak}^{meas}$. We now read off the values of $\left|S_{21}^{meas}(\omega_{m-th\,Peak}^{meas})\right|$ and determine, for all values of m, the quantities

$$Q_m = \left|S_{21}^{meas}(\omega_m)\right| - f_{centre}(\omega_m)$$

and sort these into sets of negative values Q_m^- and positive values Q_m^+. We then find the largest absolute values of these quantities as

$$\delta_{UP} = Max\{Q_m^+, \forall m\}$$

and

$$\delta_{DOWN} = Max\{|Q_m^-|, \forall m\}$$

This gives us the functions

$$f_{upper}(\omega) = f_{centre}(\omega) + \delta_{UP}$$

and

$$f_{lower}(\omega) = f_{centre}(\omega) + \delta_{DOWN}$$

also illustrated in Fig.4.5-10.

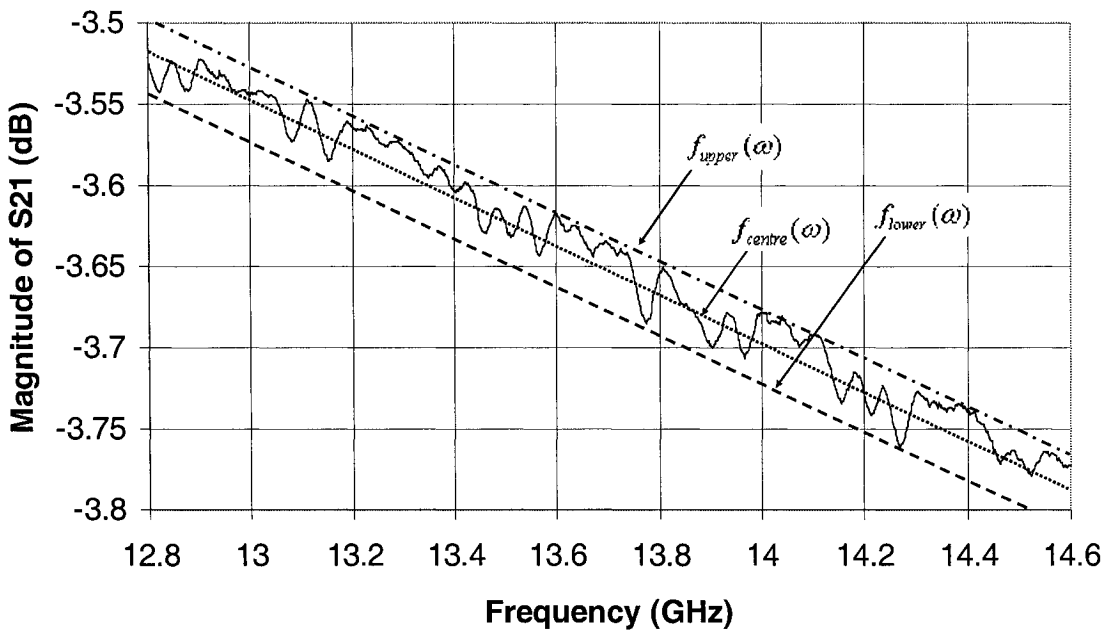


Figure 4.5-10 : Sketch to Illustrate Bounding Functions

Finally, we form the objective function

$$F_{obj}\{\alpha_o\} = \sum_{n=1}^{N} \left| f_{upper}^{meas}(\omega) + f_{lower}^{meas}(\omega) - 2 f_{centre}^{model}(\omega, \alpha_o) \right|$$

where

$$f_{centre}^{model}(\omega) = \left| S_{21}^{Model}\left\{\omega, C_1, C_2, L_1, \alpha_o = ?, \alpha_s, \ell\sqrt{\varepsilon_{eff}}\right\} \right| = \left| S_{21}^{Model}\{\omega, \alpha_o\} \right|$$

all the already-retrieved equivalent circuit parameters being used. Minimisation of $F_{obj}\{\alpha_o\}$ retrieves the value of α_o and provides Fig.4.5-11 through 4.5-14.

4.5.8 Evaluation of the Effectiveness of the Retrieval Algorithm

At this stage we have retrieved the equivalent circuit model of a wound-up length of a specific coaxial cable destined for a cable-wrap assembly, measured at ambient temperature. We have used magnitude-only measurements in the retrieval of this equivalent circuit. We now wish to use this retrieved equivalent circuit to predict the group-delay and group-delay-slope of the cable wrap, and then compare it to the measured group-delay and group-delay-slope.

Whether using the measured phase or the retrieved phase (i.e. the phase retrieved from the retrieved equivalent circuit) it is necessary to numerically differentiate the phase response $\phi(\omega) = \arg\{S_{21}(\omega)\}$ in order to find the group-delay

$$\tau_g(f) = -\frac{1}{2\pi}\frac{\partial\phi(f)}{\partial f}$$

and the group-delay slope

$$\tau_g^{slope} = \frac{\partial \tau_g(f)}{\partial f}$$

This numerical differentiation has to be performed either by software in the VNA itself or using off-line software; we will use the latter option. The group delay values at the sample frequencies can then be found using a variety of methods of estimating the numerical derivative of $\phi(\omega)$. An M-th order generalised difference estimator has been given by Boashash [35] as

$$\frac{\partial \phi(f)}{\partial f} = \sum_{m=-M/2}^{m=M/2} b_m \phi(f_{n+m})$$

with the coefficients b_m are given in [35]. We have used a fourth-order estimator

$$\left.\frac{\partial \phi(f)}{\partial f}\right|_{f=f_n} \cong \frac{\phi(f_{n-2}) - 8\phi(f_{n-1}) + 8\phi(f_{n+1}) - \phi(f_{n+2})}{12\Delta f}$$

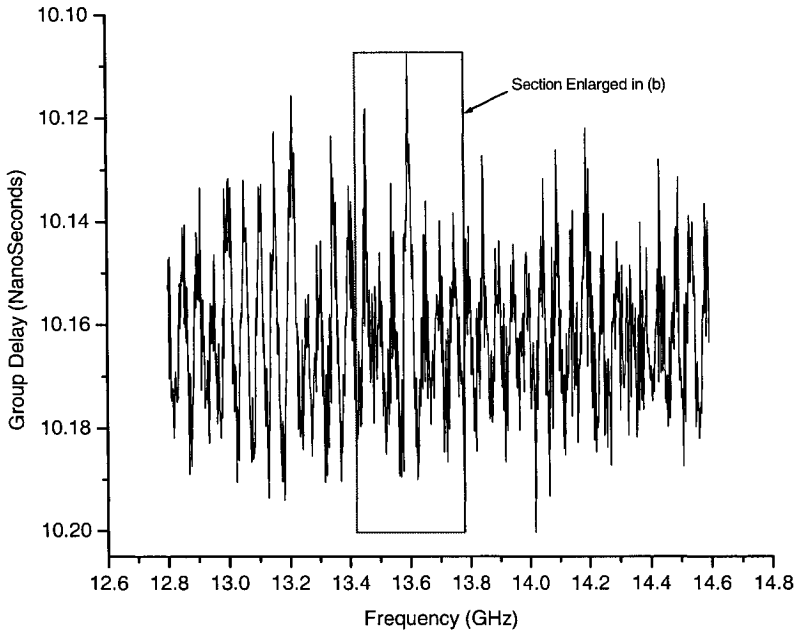
In essence, this "estimator" uses a five-point numerical derivative. The order of magnitude of the error is $O\{(\Delta f)^4\}$.

The measured phase data is noisy, and this noise will be emphasized by the numerical differentiation process. Hence, before applying numerical differentiation we first smooth measured phase data using the least-squares smoother [36,pp.570]

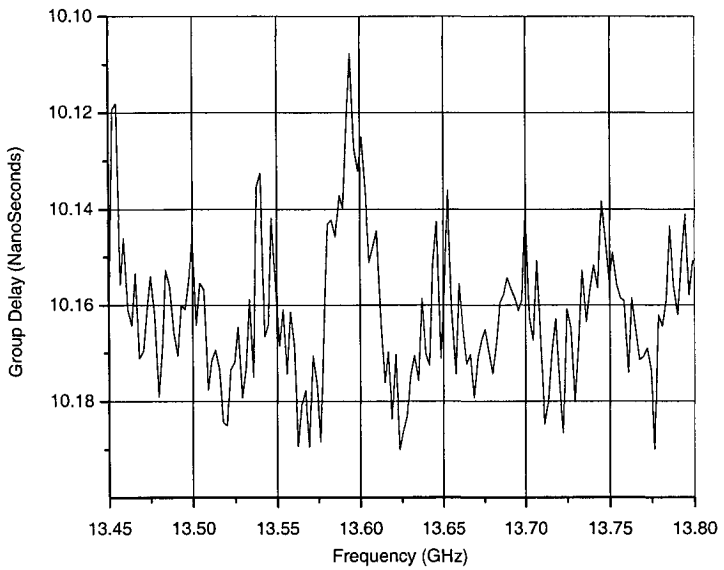
$$\phi^{Smoothed}(f_n) \cong \frac{-3\phi(f_{n-2}) + 12\phi(f_{n-1}) + 17\phi(f_n) + 12\phi(f_{n+1}) - 3\phi(f_{n+2})}{35}$$

If this is applied once to a set of data it is referred to as "single-pass" smoothing. If it is applied three times in succession it is "triple-pass" smoothing. We illustrate the effects of the smoothing operations in Figures 4.5-11 through 4.5-14. It is difficult to compare plots

which show the said quantities over the complete frequency band. In each of the above figures we have therefore included, as part (b), an enlarged plot of only a portion of the frequency band. It is clear from such parts (b) that the smoothing operation does not "falsify" the data. Note that due to the deleterious effects of differentiation we use triple-pass smoothing on the group-delay data (obtained from numerical differentiation of the phase data) to et rid of the "spikyness" before numerically differentiating it to obtain the group-delay slope.

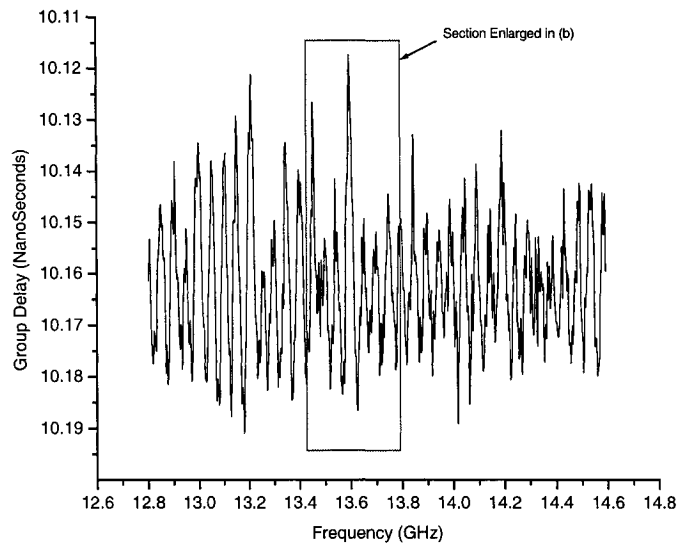


(a)

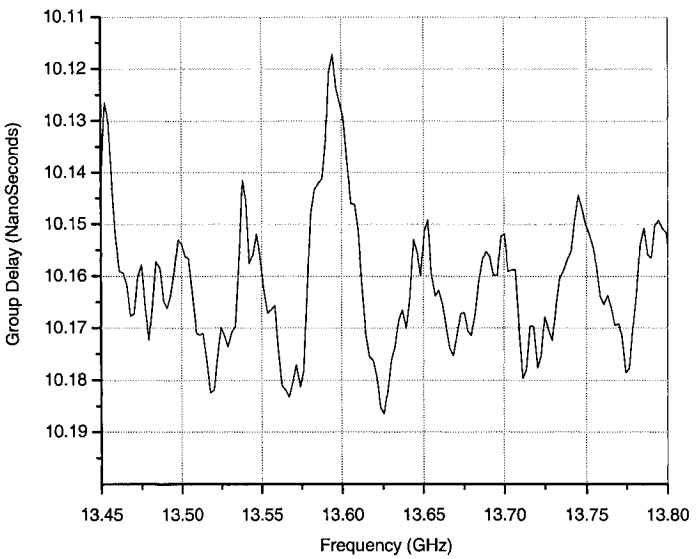


(b)

Figure 4.5-11 : Group-Delay Using Non-Smoothed Measured Phase
(a). Data Over Complete Frequency Range
(b). Data Over Portion of Frequency Range Indicated in (a).



(a)



(b)

Figure 4.5-12 : Group-Delay Using Smoothed (Single-Pass) Measured Phase
(a). Data Over Complete Frequency Range
(b). Data Over Portion of Frequency Range Indicated in (a).

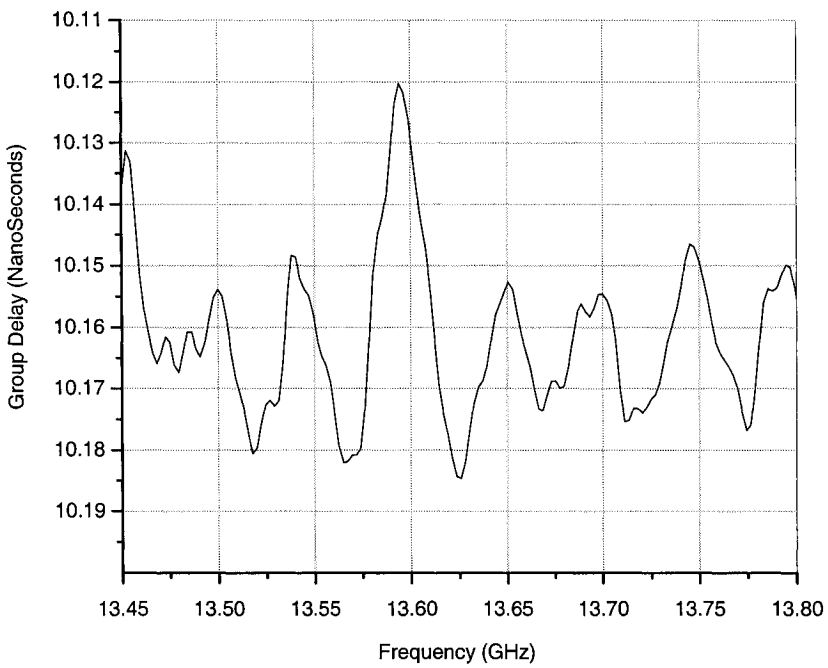
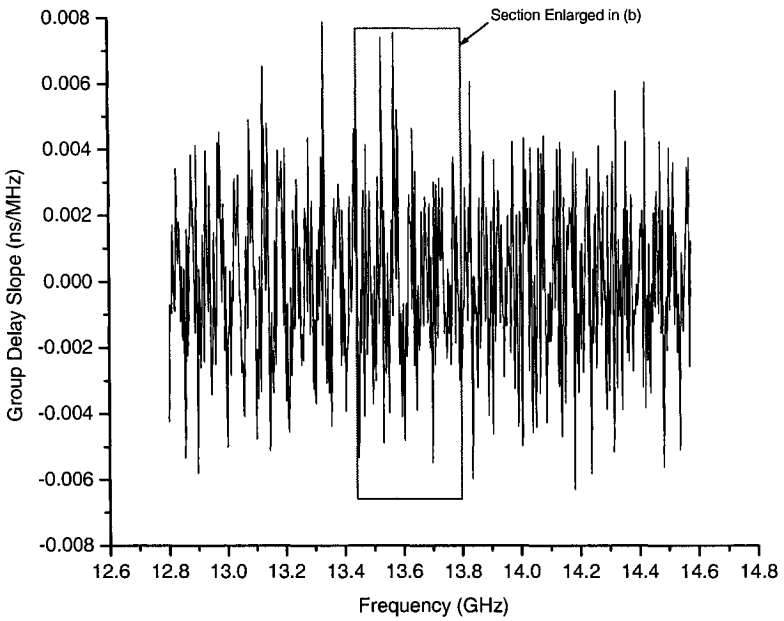
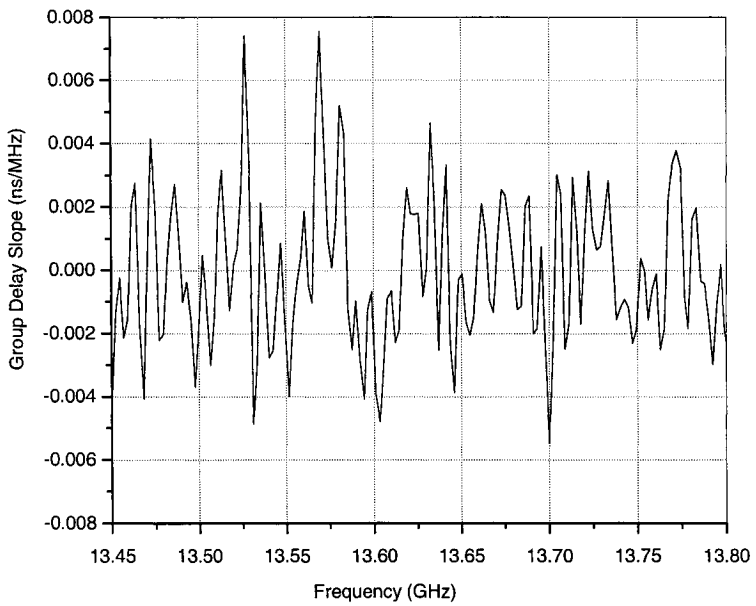


Figure 4.5-13 : Group-Delay from Figure 4.5-12 (b) After Triple-Pass Smoothing

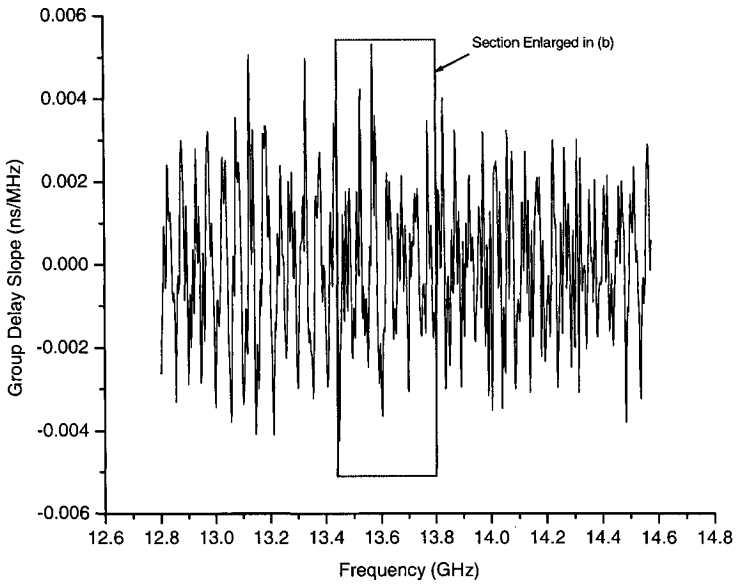


(a)

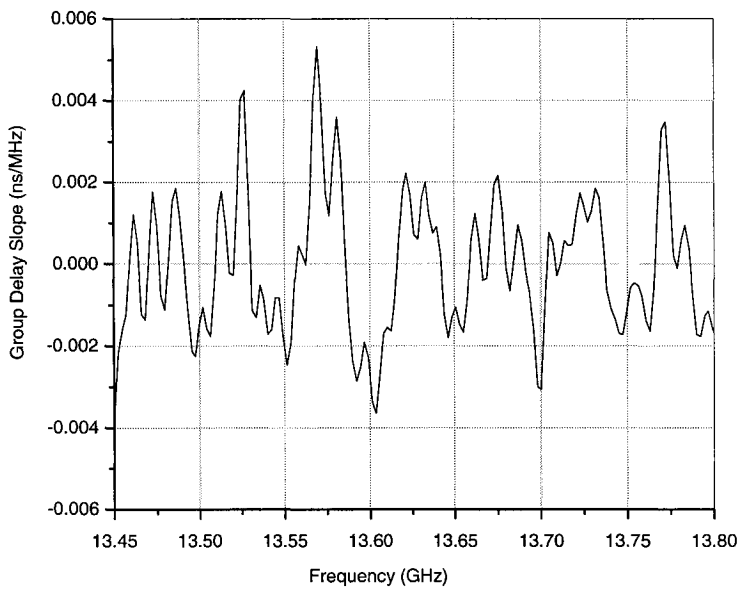


(b)

Figure 4.5-14 : Group-Delay-Slope Determined Using Non-Smoothed Group-Delay
(a). Data Over Complete Frequency Range
(b). Data Over Portion of Frequency Range Indicated in (a).



(a)



(b)

Figure 4.5-15 : Group-Delay-Slope Using Smoothed (Triple-Pass) Group-Delay
(a). Data Over Complete Frequency Range
(b). Data Over Portion of Frequency Range Indicated in (a).

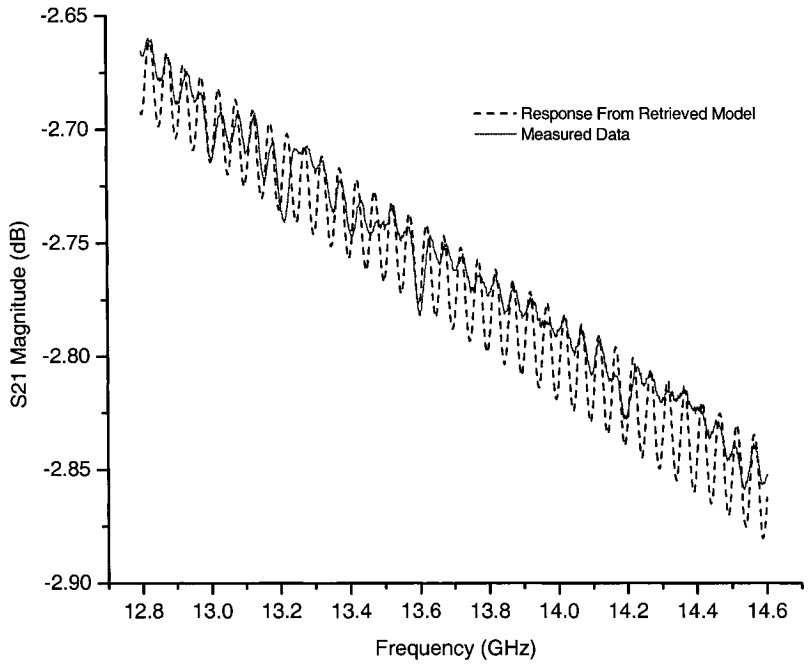
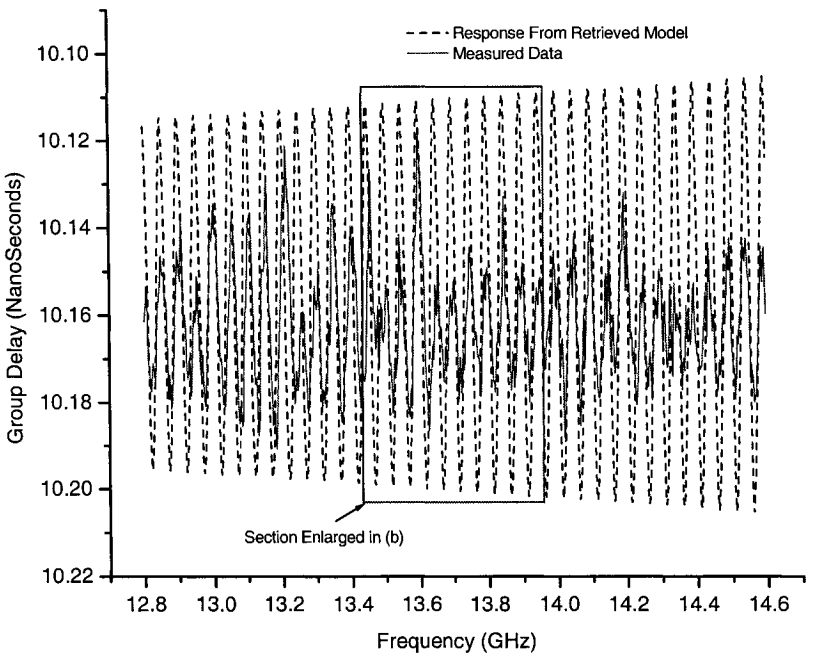
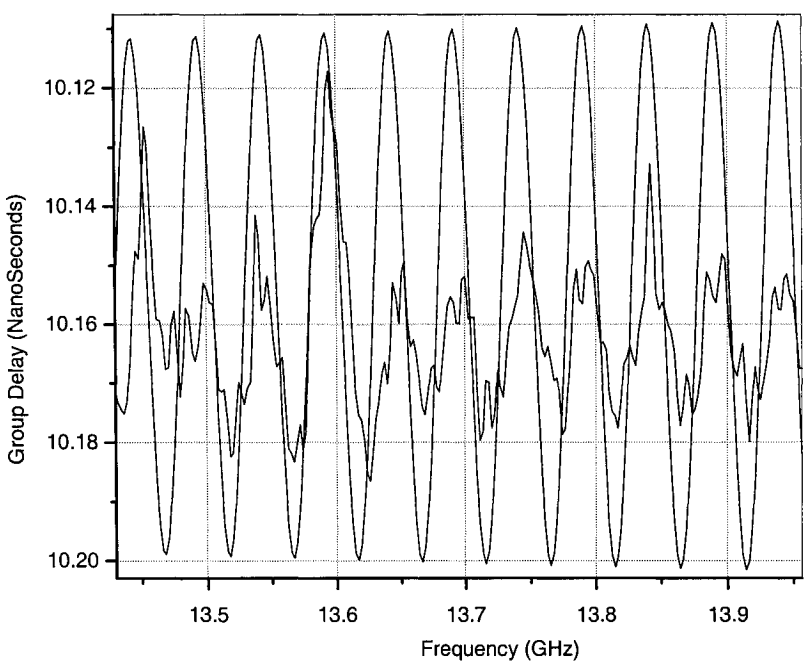


Figure 4.5-16 : $|S_{21}|$ from Retrieved Equivalent Circuit Compared to Actual Measured $|S_{21}|$



(a)



(b)

Figure 4.5-17 : Retrieved Group-Delay Compared to Actual Measured Group-Delay

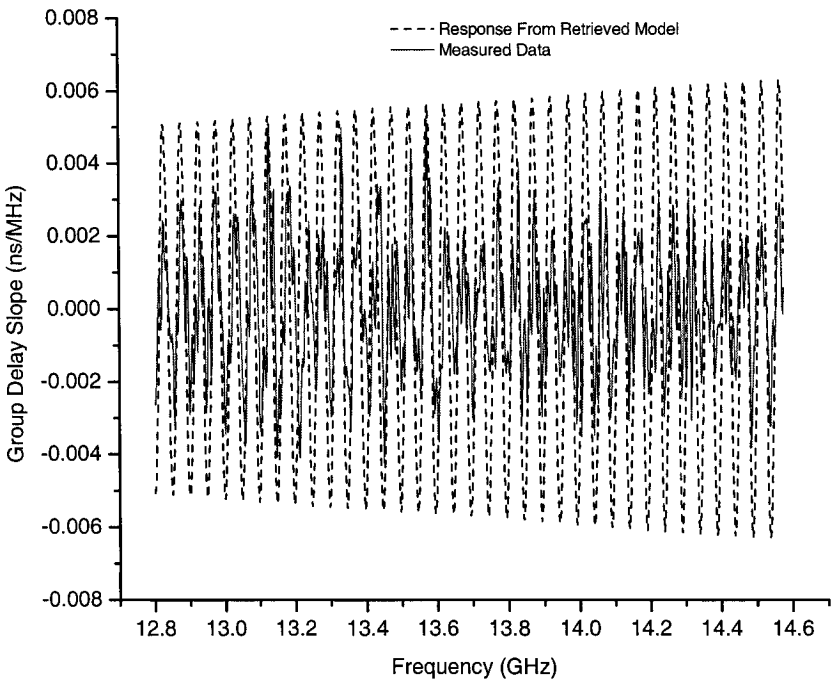


Figure 4.5-18 : Retrieved Group Delay Slope Compared to Actual Measured Group-Delay Slope

4.5.9 Establishment of the Retrieval Algorithm Through Further Examples

The same cable-wrap was re-measured, over the same frequency band, but at a temperature far below 0°C (the "cold" case) and at a temperature slightly below 100°C (the "hot" case). In each case the magnitude only data is used to retrieve the equivalent circuit using the ambient equivalent circuit values as starting values. In the graphed data that follows we show the retrieved group-delay and group-delay slope in each case, compared to the measured data for these quantities obtained from vector measurements.

A. Cable-Wrap at Low Temperature

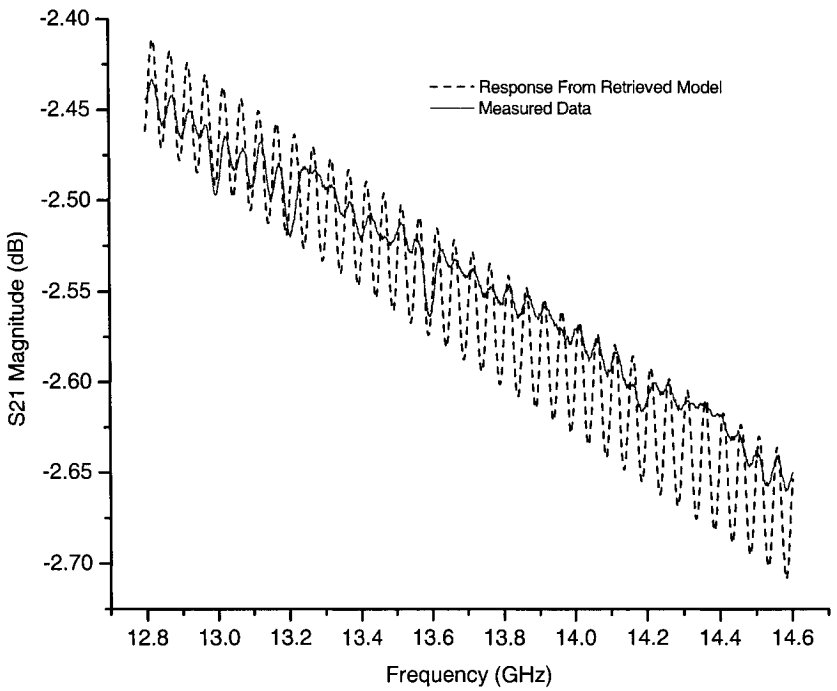


Figure 4.5-19 : $|S_{21}|$ from Retrieved Equivalent Circuit Compared to Actual Measured $|S_{21}|$

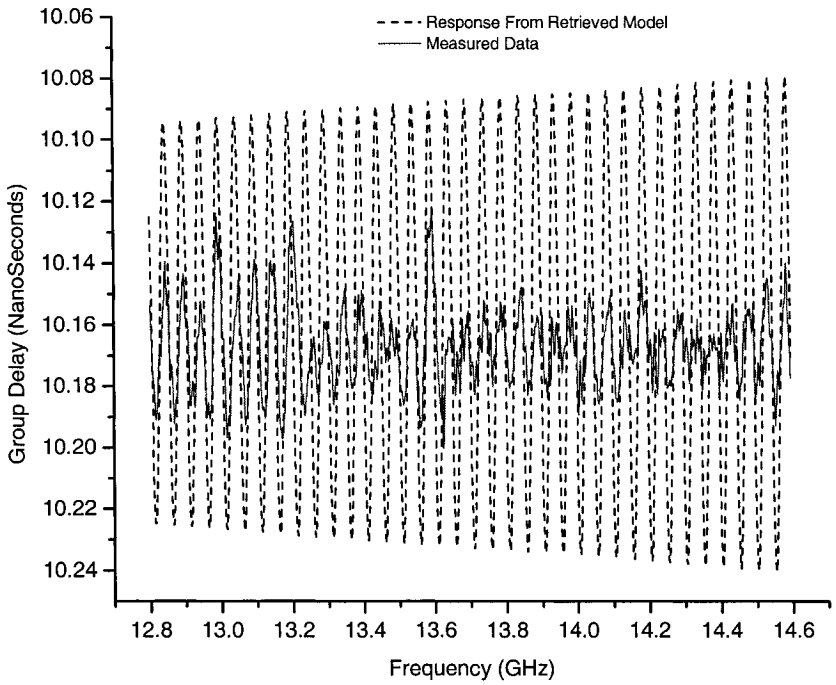


Figure 4.5-20 : Retrieved Group-Delay Compared to Actual Measured Group-Delay

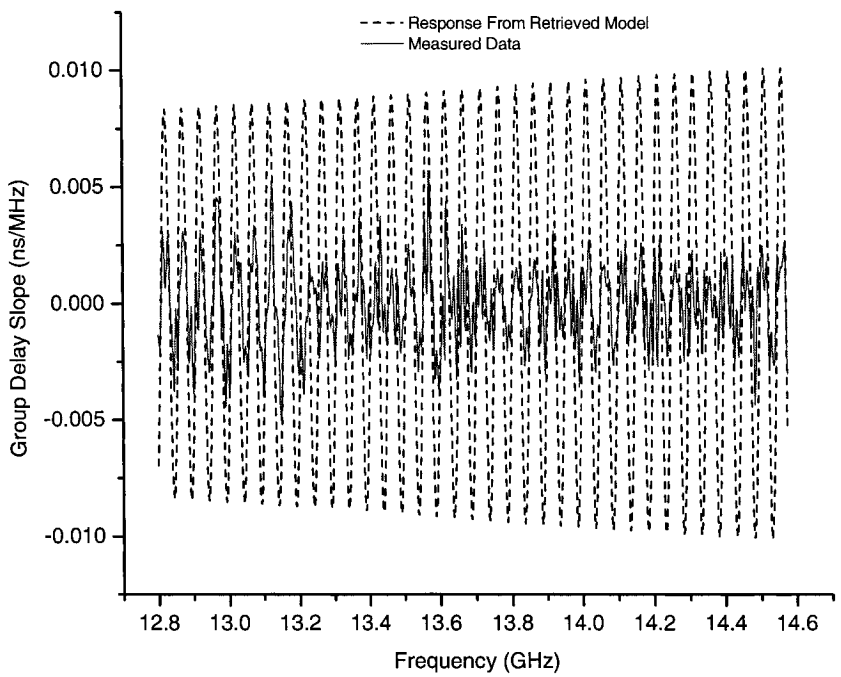


Figure 4.5-21 : Retrieved Group Delay Slope Compared to Actual Measured Group-Delay Slope

B. Cable-Wrap at High Temperature

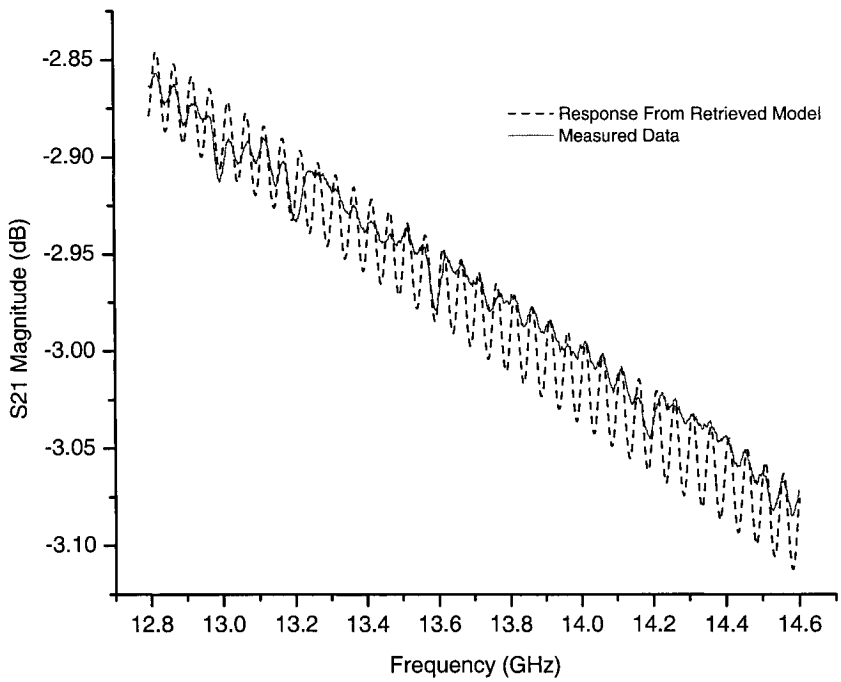


Figure 4.5-22 : $|S_{21}|$ from Retrieved Equivalent Circuit Compared to Actual Measured $|S_{21}|$

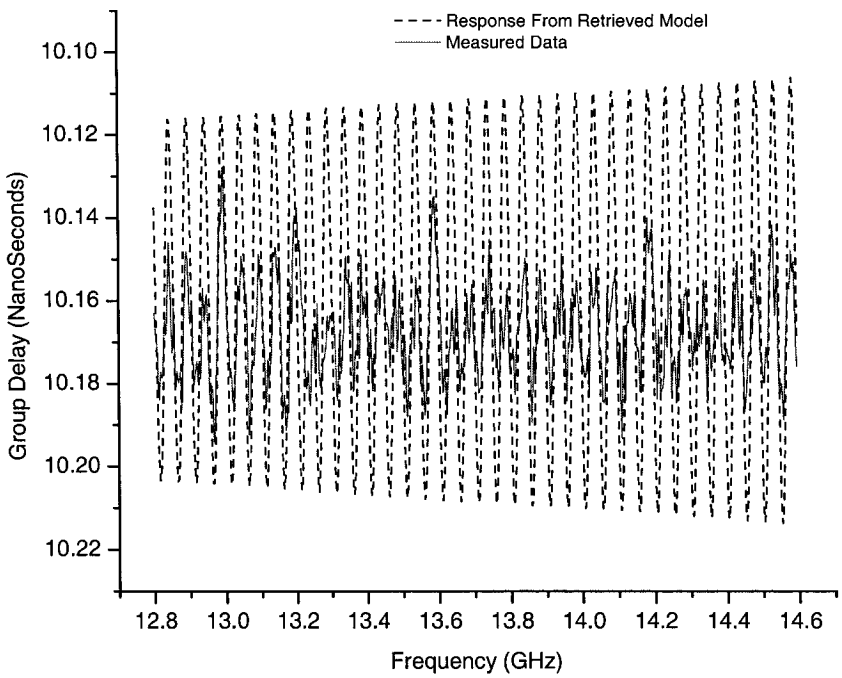


Figure 4.5-23 : Retrieved Group-Delay Compared to Actual Measured Group-Delay

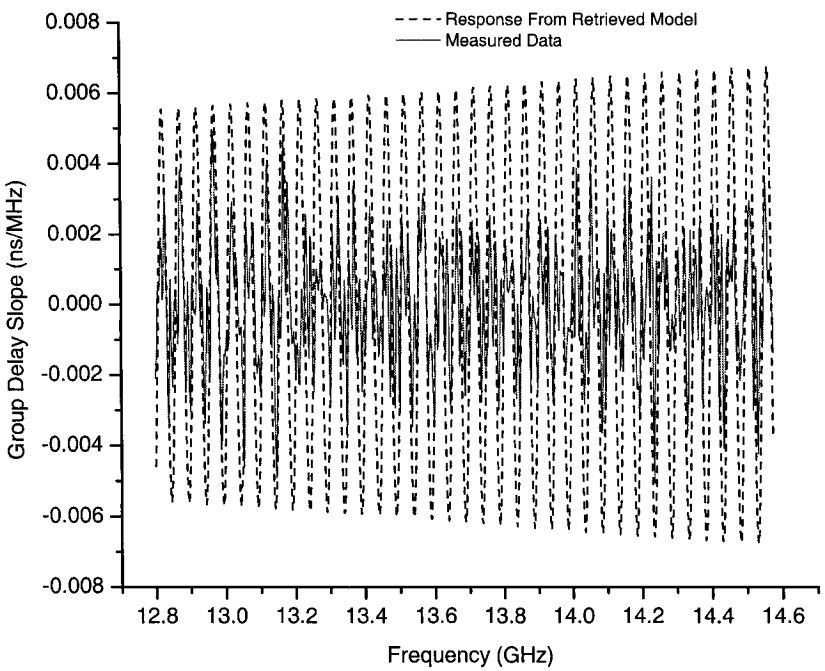


Figure 4.5-24 : Retrieved Group Delay Slope Compared to Actual Measured Group-Delay Slope

4.6 CONCLUSIONS

We have developed a new approach, which we have called model-based phase retrieval, that utilises the known design knowledge of particular devices in order to allow the unique retrieval of phase information from measured magnitude information. The proposed method has previously not been suggested elsewhere.

4.7 REFERENCES FOR CHAPTER 4

[1] R.Levy, "Derivation of equivalent circuits of microwave structures using numerical techniques", IEEE Trans. Microwave Theory Tech., Vol.47, No.9, pp.1688-1695, Sept.1999.

[2] R.Levy, "Determination of simple equivalent circuits of interacting discontinuities in waveguides or transmission lines", IEEE Trans. Microwave Theory Tech., Vol.48, No.10, pp.1712-1716, Oct.2000.

[3] T.Mangold & P.Russer, "Full-wave modeling and automatic equivalent-circuit generation of millimeter-wave planar and multilayer structures", IEEE Trans. Microwave Theory Tech., Vol.47, No.6, pp.851-858, June 1999.

[4] I.Timmins & K-L. Wu, "An efficient systematic approach to model extraction for passive microwave circuits", IEEE Trans. Microwave Theory Tech., Vol.48, No.9, pp.1565-1573, Sept.2000.

[5] Q.J.Zhang & K.C.Gupta, "Neural Networks for RF and Microwave Design" (Artech House Inc., 2000).

[6] F.Wang & Q.J.Zhang, "Knowledge-based neural models for microwave design", IEEE Trans. Microwave Theory Tech., Vol.45, pp.2333-2343, Sept.1997.

[7] X.Ding, V.K.Devabhaktuni, B.Chattaraj, M.C.E.Yagoub, M.eo, J.Xu & Q.J.Zhang, "Neural-network approaches to electromagnetic-based modeling of passive components and their applications to high-frequency and high-speed nonlinear circuit optimization", IEEE Trans. Microwave Theory Tech., Vol.52, No.1, pp.436-449, Jan.2004

[8] J.N.Brittingham, E.K.Miller & J.L.Willows, "Pole extraction from real-frequency information", Proc.IEEE, Vol.68, No.2, pp.263-273, Feb.1980.

[9] G.J.Burke, E.K.Miller, S.Chakrabarti & K.Demarest, "Using model-based parameter estimation to increase the efficiency of computing electromagnetic transfer function", IEEE Trans. Magnetics, Vol.25, No.4, pp.2807-2809, July 1989.

[10] E.K.Miller & G.J.Burke, "Using model-based parameter estimation to increase the physical interpretability and numerical efficiency of computational electromagnetics", Computer Physics Communications, Vol.68, pp.43-75, Nov.1991.

[11] K.Kottapalli, T.K.Sarkar, Y.Hua, E.K.Miller & G.J.Burke, "Accurate calculation of wide-band response of electromagnetic systems utilising narrow-band information", IEEE Trans. Microwave Theory Tech., Vol.39, No.4, pp.682-687, April 1991.

[12] K.Kottapalli, T.K.Sarkar, Y.Hua, E.K.Miller & G.J.Burke, "Accurate calculation of wide-band response of electromagnetic systems utilising narrow-band information", Computer Physics Communications, Vol.68, pp.126-144, Nov.1991.

[13] R.S.Adve & T.K.Sarkar, "Generation of accurate broadband information from narrowband data using the Cauchy method", Microwave & Optical Tech. Letters, Vol.6, No.10, pp.569-573, Aug.1993.

[14] R.S.Adve & T.K.Sarkar, "The effect of noise in the data on the Cauchy method", Microwave & Optical Tech. Letters, Vol.7, No.5, pp.242-247, April 1994.

[15] T.Dhaene, J.Ureel, N.Faché & D. de Zutter, "Adaptive frequency sampling algorithm for fast and accurate S-parameter modeling of general planar structures", IEEE MTT-S Int. Symposium Digest, pp.1427-1430, 1995.

[16] R.S.Adve, T.K.Sarkar, S.M.Rao, E.K.Miller & D.R.Pflug, "Application of the Cauchy method for extrapolating/interpolating narrow-band system responses", IEEE Trans. Microwave Theory Tech., Vol.45, No.5, pp.837-845, May 1997.

[17] E.K.Miller, "Model-based parameter estimation in electromagnetics : Part I – Background and theoretical development", IEEE Antennas & Propagation Magazine, Vol.40, pp.42-52, Feb.1998.

[18] E.K.Miller, "Model-based parameter estimation in electromagnetics : Part II – Applications to EM Observables", IEEE Antennas & Propagation Magazine, Vol.40, pp.51-65, April 1998.

[19] J.E.Bracken, D.Sun & Z.Cendes, "S-domain methods for simultaneous time and frequency characterisation of electromagnetic devices", IEEE Trans. Microwave Theory Tech., Vol.46, No.9, pp.1277-1290, Sept.1998.

[20] A.G.Lampérez, T.K.Sarkar & M.Salazar-Palma, "Generation of accurate rational models of lossy systems using the Cauchy method", IEEE Microwave & Wireless Components Letters, Vol.14, No.10, pp.490-492, Oct.2004.

[21] T.K.Sarkar & O.Pereira, "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials", IEEE Antennas & Propagation Magazine, Vol.37, No.1, pp.48-55, Feb.1995.

[22] IMSL Math Library, Routine *BDCPOL*.

[23] W.H.Swann, "Constrained optimisation by direct search" in P.E.Gill & W.Murray (Edits.) : *Numerical Methods for Constrained Optimization* (Academic Press, 1974 pp.191-217.

[24] I.Bahl, *Lumped Elements for RF and Microwave Circuits* (Artech House, 2003).

[25] M.A.Ismail, D.Smith, A.Panariello, Y.Wang & M.Yu, "EM-based design of large-scale dielectric-resonator filters and multiplexers by space mapping", IEEE Trans. Microwave Theory Tech., Vol.MTT-52, No.1, pp.386-392, Jan.2004.

[26] A.E.Atia & A.E.Williams, "Narrow-bandpass waveguide filters", IEEE Trans. Microwave Theory Tech., Vol.MTT-20, pp.258-265, April 1972.

[27] J.D.Rhodes, *Theory of Electrical Filters* (Wiley, 1976).

[28] S.Amari, "Synthesis of cross-coupled resonator filters using an analytical gradient-based optimisation technique", IEEE Trans. Microwave Theory Tech., Vol.48, pp.1559-1564, Sept.2000.

[29] D.Pozar, *Microwave Engineering* (Wiley, 2004).

[30] M.Yu, Private Communication.

[31] F.Croq, F.Dolmeta, B.Lejay, E.Vourch & M.Reynaud, "Mechanically scanned, electronically zoomed, receive and transmit antennas for low Earth orbit spaceborne wideband applications in Ku band", IEEE AP-S Symp. Digest, pp.224-227, 2002.

[32] M.L.Majewski, R.W.Rose & J.R.Scott, "Modeling and characterization of microstrip-to-coaxial transitions", IEEE Trans. Microwave Theory Tech., Vol.MTT-29, No.8, pp.799-805, August 1981.

[33] J.R.Juroshek, "A study of measurements of connector repeatability using highly reflecting loads", IEEE Trans. Microwave Theory Tech., Vol.MTT-35, No.4, pp.457-460, April 1987.

[34] H.A.Haus, *Waves and Fields in Optoelectronics* (Prentice-Hall, 1984).

[35] B.Boashash, "Estimating and interpreting the instantaneous frequency of a signal – Part 2 : Algorithms and applications", Proc. IEEE, Vol.80, No.4, pp.540-568, April 1992.

[36] R.W.Hamming, *Numerical Methods for Scientists and Engineers* (Dover Publ., 1973).

[37] Y.C.Shih & T.Itoh, "Transmission Lines and Waveguides" in: Y.T.Lo & S.W.Lee (Edits.), *Antenna Handbook, Vol.IV* (Van Nostrand Reinhold, 1993) Chap.28.

Chapter 5

General Conclusions

The contributions of this thesis are as follows:

- We have proposed a new method for the retrieval of phase-related quantities from magnitude-only measurements on two-port microwave systems. The method is applicable when a microwave system is manufactured in relatively large quantities. A reduced-order model is developed using the prototype of such a system. Such a reduced-order model will include a number of quantities or "coefficients" (other than frequency) whose value is determined using a full vector data set on the prototype. Establishing the form required for the reduce-order model acts as the additional constraints that are needed to provide a unique relationship between the transfer function magnitude and its phase. When later production items are tested using magnitude-only measurements, optimization is used (with the above-mentioned "coefficients" as the optimization variables, for which only small adjustments are needed from one device to the next) to match the measured and modelled magnitude responses. The retrieved "coefficients" are then used to predict the phase response using the model (that is, to retrieve the phase response).
- The use of the model-based retrieval technique has been demonstrated through its application to three different types of "device". The first is a bridged-T network, selected for its "obvious" non-mimimum phase properties. In this example the "measured" data was artificially generated so as to simulate a prototype item and

subsequent production line items. A rational function representation was used as the model in the model-based phase-retrieval. The second example used actual measured data of a microwave filter that, because of the cross-coupling, is also non-mininum phase network. In this example an equivalent circuit representation was used as the model in the model-based phase-retrieval. Finally, a cable-wrap system was considered, for which an equivalent circuit representation was also used as the model in the model-based phase-retrieval procedure. This represents a class of problems for which the networks are electrically long, have a rapid "ripple" in their response as a function of frequency, but have a transmission coefficient always relatively close to unity. In such cases it is the group delay variation and group delay slope are of principal interest, and we have shown how customized model-based phase-retrieval approaches may be developed to suit such special cases. Thus it has been demonstrated that a range of reduced-order model types can be utilized in proposed retrieval technique.

Appendix I

Computation of the Scattering Parameters for the Cable-Wrap Assembly Equivalent Circuit Used in Section 4.5

We here summarise the circuit analysis of the equivalent circuit shown in Figure 4.5-1.

The procedure is as follows:

- Determine the ABCD-matrix of each of the five sections.
- Multiply these five ABCD matrices.
- Convert the ABCD parameters to S-parameters.

We first list the ABCD parameters of each of the five sections in Figure 4.5-1[1] :

- ABCD parameters of the short straight coaxial section of length l_{in}

$$A = \cos(\beta_{con} l_{in}) \qquad B = jZ_{con} \sin(\beta_{con} l_{in})$$

$$C = jY_{con} \sin(\beta_{con} l_{in}) \qquad D = \cos(\beta_{con} l_{in})$$

- ABCD parameters of the discontinuity region on the left:

$$A = 1 + \frac{Y_b}{Y_c} \qquad B = \frac{1}{Y_c} \qquad C = Y_a + Y_b + \frac{Y_a Y_b}{Y_c} \qquad D = 1 + \frac{Y_a}{Y_c}$$

$$Y_a = j\omega C_{11} \qquad Y_b = j\omega C_{12} \qquad Y_c = \frac{1}{j\omega L_1}$$

- ABCD parameters of the flexible coaxial cable

$$A = \cos(\beta l) \qquad B = jZ_0 \sin(\beta l)$$

$$C = jY_0 \sin(\beta l) \qquad D = \cos(\beta l)$$

- ABCD parameters of the discontinuity region on the right:

$$A = 1 + \frac{Y_b}{Y_c} \qquad B = \frac{1}{Y_c} \qquad C = Y_a + Y_b + \frac{Y_a Y_b}{Y_c} \qquad D = 1 + \frac{Y_a}{Y_c}$$

$$Y_a = j\omega C_{22} \qquad Y_b = j\omega C_{21} \qquad Y_c = \frac{1}{j\omega L_2}$$

- ABCD parameters of the short straight coaxial section of length l_{out}

$$A = \cos(\beta_{con} l_{out}) \qquad B = jZ_{con} \sin(\beta_{con} l_{out})$$

$$C = jY_{con} \sin(\beta_{con} l_{out}) \qquad D = \cos(\beta_{con} l_{out})$$

Once the parameters A_T, B_T, C_T and D_T of the complete circuit have been obtained through matrix multiplication (at each frequency ω), the S-parameters are then found as [1]:

$$S_{11} = \frac{A_T + B_T / Z_0 - C_T Z_0 - D_T}{A_T + B_T / Z_0 + C_T Z_0 + D_T}$$

$$S_{12} = S_{21}$$

$$S_{21} = \frac{2}{A_T + B_T / Z_0 + C_T Z_0 + D_T}$$

$$S_{22} = \frac{-A_T + B_T / Z_0 - C_T Z_0 + D_T}{A_T + B_T / Z_0 + C_T Z_0 + D_T}$$

For the analysis of the "reduced" equivalent circuit in Figure 4.5-4 we simply omit the ABCD matrices of the short straight coaxial sections at the beginning and end.

References for Appendix I

[1] D.M. Pozar, *Microwave Engineering* (Wiley,2005) Third Edition.

Appendix II

Computation of the Scattering Parameters for the Bridged-T Circuit Used in Section 4.3

We wish to detail the analysis of the bridged-T 2-port network shown below in Fig.II.1, and used in Section 4.4.

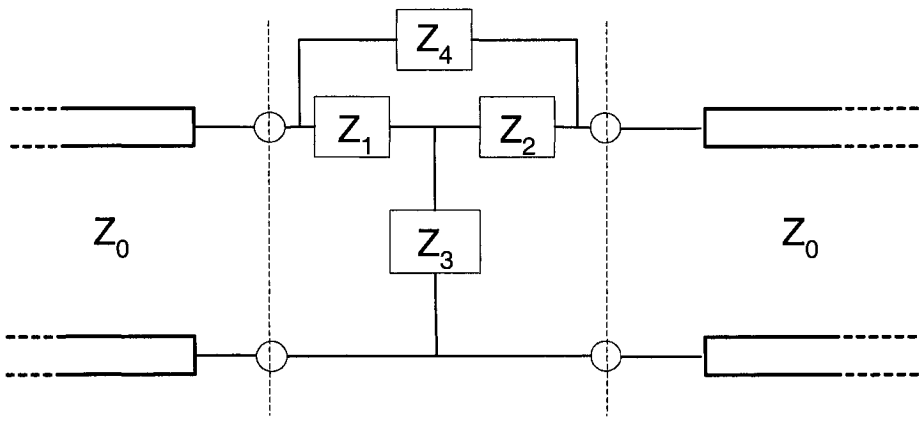


Figure II.1 : Bridged-T Two-Port Network

The admittance matrix of this 2-port network can be shown, using information given in Weinberg [1,pp.75-77] , to be

$$[Y]=\begin{bmatrix} \frac{Z_2+Z_3}{\Delta Z}+\frac{1}{Z_4} & \frac{-Z_3}{\Delta Z}-\frac{1}{Z_4} \\ \frac{-Z_3}{\Delta Z}-\frac{1}{Z_4} & \frac{Z_1+Z_3}{\Delta Z}+\frac{1}{Z_4} \end{bmatrix}$$

where

$$\Delta Z = Z_1Z_2 + Z_1Z_3 + Z_2Z_3$$

The conversion from the admittance parameters to the scattering parameters is then [2,pp.187] simply

$$S_{11} = \frac{(Y_0 - Y_{11})(Y_0 + Y_{22}) + Y_{12}Y_{21}}{\Delta Y}$$

$$S_{12} = \frac{-2Y_0Y_{12}}{\Delta Y}$$

$$S_{12} = \frac{-2Y_0Y_{21}}{\Delta Y}$$

$$S_{22} = \frac{(Y_0 + Y_{11})(Y_0 - Y_{22}) + Y_{12}Y_{21}}{\Delta Y}$$

where

$$\Delta Y = (Y_0 + Y_{11})(Y_0 + Y_{22}) - Y_{12}Y_{21}$$

References for Appendix II

[1] L.Weinberg, *Network Analysis and Synthesis* (McGraw-Hill, 1962).

[2] D.M.Pozar, Microwave Engineering (Wiley,2005) Third Edition.