DCU

Transient Simulation of Complex Electronic Circuits and Systems Operating at Ultra High Frequencies

by

Emira Dautbegović

A dissertation submitted in fulfilment of the
requirements for the award of

Doctor of Philosophy

to

Dublin City University
School of Electronic Engineering

Supervisor: Dr. Marissa Condon

May 2005

DECLARATION

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed Dautbegovic Emina

Student ID 51169487

Date 12th May, 2005

ACKNOWLEDGMENTS

This thesis could never be written without continuous support, understanding and love from my wonderful parents Dževad and Hidajeta. From the bottom of my heart THANK YOU for being there for me!
Ova teza ne bi nikada mogla biti napisana bez stalne potpore, razumjevanja i ljubavi mojih roditelja Dževada i Hidajete. Iz dubine moga srca HVALA vam što ste bili uz mene!

I would like to express my sincere gratitude and appreciation to my supervisor, Dr. Marissa Condon, for her guidance and much needed encouragement. Her motivation, helpful suggestions and patience made the journey through this thesis an unforgettable experience. Thanks a million for all your help!

I have met many colleagues here at Dublin City University that made my time here such enjoyable experience. A very special thanks goes to Orla. Moving to Ireland to embark on uncertain journey called Ph.D, I could not have dreamed that I would meet such unique person that was nothing less than a sister to me. I would like to thank Lejla for being such a good friend to me and making me laugh in the most impossible situations. I will be always grateful for that "homely feeling" to the small Bosnian community here at DCU: Dalen for all his support, especially in the early days of my Irish adventure; Amra for drawing my attention to the research opportunity that became the corner stone in my pursuit towards a PhD; and Damir for his very useful insights in the art of writing dissertation. Thanks to Es for being such a wonderful housemate. For all the good times in and out of the lab the "guilty" people are Li-Chuan (thanks for cooking me decent dinners on so many occasions), Tam, Javier, Rossen, John and Aubrey. I am grateful to Aida, Jasmina, Adis, Kevin and Armin for their friendship and support. I had great time while serving in the DCU Postgraduate Society Committee as Sports officer; those basketball and volleyball games are going to remain a very nice memory for me. Thanks to the all eager judokas, especially to Nicolas, Grainne and Maeve, for fun on and off the mat. I am proud to call all of you guys my FRIENDS. Thank you all for the good times I had here in Ireland!

There are several people in Sarajevo to whom I would like to thank for all their support and friendship during my Irish adventure. A special thanks goes to my family: *Vedo, Dedo, Beco, Daja, Aida, Jasmina, Edina i Nedime hvala vam na svoj vašoj podršci.* I am grateful to Arijana, Dženita and Nermina for being such true friends. I am thankful to my professors and colleagues at the Faculty of Electrical Engineering, University of Sarajevo for stimulating my interest in control and electronic engineering.

Finally, thanks goes to the IBM and Synopsis, who awarded me their fellowships and thus provided much welcome funding towards this research. I am thankful to Mr. Jim Dowling and Dr. Connor Brennan, and other employees of the School of Electrical Engineering at Dublin City University for their support during my study. I am grateful to Prof. Thomas Brazil and the University College Dublin Microwave Research Group for providing circuit examples and measured data for testing purposes.

I would like to acknowledge all the exceptional people who made my Ph.D. study such a joyful experience. I am sure that there are a number of people who I failed to mention here. This is not due to a lack of gratefulness, just poor memory…

This thesis is dedicated to my beloved family.
Ovu tezu posvećujem mojoj voljenoj porodici.

Transient Simulation of Complex Electronic Circuits and Systems Operating at Ultra High Frequencies

Emira Dautbegović

ABSTRACT

The electronics industry worldwide faces increasingly difficult challenges in a bid to produce ultra-fast, reliable and inexpensive electronic devices. Electronic manufacturers rely on the Electronic Design Automation (EDA) industry to produce consistent Computer Aided Design (CAD) simulation tools that will enable the design of new high-performance integrated circuits (IC), the key component of a modern electronic device. However, the continuing trend towards increasing operational frequencies and shrinking device sizes raises the question of the capability of existing circuit simulators to accurately and efficiently estimate circuit behaviour.

The principle objective of this thesis is to advance the state-of-art in the transient simulation of complex electronic circuits and systems operating at ultra high frequencies. Given a set of excitations and initial conditions, the research problem involves the determination of the transient response of a high-frequency complex electronic system consisting of linear (interconnects) and non-linear (discrete elements) parts with greatly improved efficiency compared to existing methods and with the potential for very high accuracy in a way that permits an effective trade-off between accuracy and computational complexity.

High-frequency interconnect effects are a major cause of the signal degradation encountered by a signal propagating through linear interconnect networks in the modern IC. Therefore, the development of an interconnect model that can accurately and efficiently take into account frequency-dependent parameters of modern non-uniform interconnect is of paramount importance for state-of-art circuit simulators. Analytical models and models based on a set of tabulated data are investigated in this thesis. Two novel, highly accurate and efficient interconnect simulation techniques are developed. These techniques combine model order reduction methods with either an analytical resonant model or an interconnect model generated from frequency-dependent *s*-parameters derived from measurements or rigorous full-wave simulation.

The latter part of the thesis is concerned with envelope simulation. The complex mixture of profoundly different analog/digital parts in a modern IC gives rise to multi-time signals, where a fast changing signal arising from the digital section is modulated by a slower-changing envelope signal related to the analog part. A transient analysis of such a circuit is in general very time-consuming. Therefore, specialised methods that take into account the multi-time nature of the signal are required. To address this issue, a novel envelope simulation technique is developed. This technique combines a wavelet-based collocation method with a multi-time approach to result in a novel simulation technique that enables the desired trade-off between the required accuracy and computational efficiency in a simple and intuitive way. Furthermore, this new technique has the potential to greatly reduce the overall design cycle.

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

GLOSSARY OF TERMS

AB	Adams-Bashfort method
AM	Adams-Moulton method
ARMA	Auto-Regressive Moving Average modelling
AWE	Asymptotic Waveform Evaluation
BDF	Backward Differentiation Formula
BE	Backward Euler
BER	Bit Error ratio
CAD	Computer Aided Design
CANCER	Computer Analysis of Nonlinear Circuits Excluding Radiation
CFH	Complex Frequency Hopping
CPU	Central Processing Unit
CWT	Continuous Wavelet Transform
DAE	Differential Algebraic Equation
DFT	Direct Fourier Transformation
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EDA	Electronic Design Automation
EEG	Electroencephalogram
EM	Electro Magnetic
EMC	Electro Magnetic Compatibility
EMI	Electro Magnetic Interference
EMRA	Exponential Matrix-Rational Approximation
FE	Forward Euler
FFT	Fast Fourier Transform
FIR	Finite Impulse Response filter
FT	Fourier Theory
HB	Harmonic Balance
HF	High Frequency
IC	Integrated Circuit
ICT	Integrated Congruence Transform
IFFT	Inverse Fast Fourier Transform
IVP	Initial Value Problem
KCL	Kirchoff's Current Law
LF	Low Frequency
LU	Lower-Upper matrix decomposition
MC	Method of Characteristics
MCM	Multi-Chip Modules
MESFET	Metal-Semiconductor-Field-Effect-Transistor
MNA	Modified Nodal Analysis
MOR	Model Order Reduction
MPDE	Multi-Time Partial Differential equation
MRA	Multiresolution analysis
MTL	Multiconductor Transmission Line
ODE	Ordinary Differential Equation
p.u.l.	Per-unit-length parameter
PCB	Printed Circuit Board
PDE	Partial Differential Equation

GLOSSARY OF TERMS (continued)

PEEC	Partial Element Equivalent Circuit
PEECr	Partial Element Equivalent Circuit with retardation
PRIMA	Passive Reduced-order Interconnect Macromodelling Algorithm
PVL	Padé Via Lanczos
RF	Radio Frequency
RK	Runge-Kutta method
RKF	Runge-Kutta Fehlberg method
RLC	Resistor-inductor-capacitor (network)
ROM	Reduced Order Model
SoC	System on Chip
SPICE	Simulation Program with Integrated Circuits Emphasis
STFT	Short Term Fourier Transform
STL	Single Transmission Line system
TEM	Transverse Electromagnetic
TF	Transfer Function
TR	Trapezoidal method
VCO	Voltage-Controlled Oscillator
VLSI	Very Large Scale Integrated circuits
WT	Wavelet Theory

LIST OF SYMBOLS

α	Attenuation constant
β	Phase constant
γ	Complex propagation constant
δ	Skin depth
$\delta(t)$	Kronecker delta function.
ε	Error tolerance
λ	Wavelength [m]
ω	Radian frequency [rad/s]
μ	Magnetic permeability
ρ	Relative electrical resistance
τ	Time constant; line delay
Δx	Length of a short piece of the line
$\varphi(x)$	Interior scaling function
$\varphi_b(x)$	Boundary scaling function
$\psi(x)$	Interior mother wavelet function
$\psi_b(x)$	Boundary mother wavelet function
$\eta(x)$	Spline function
$\Psi_k(t)$	"Wavelets"
$\Psi_x^{\psi}(\tau,s)$	Continuous wavelet transform
$\boldsymbol{A}$	A real matrix A
$\boldsymbol{A}^T$	A real matrix A transposed
$a(t)$, $b(t)$	A time dependant coefficient in series/Padé expansion
C	Capacitance [F]; also capacitance p.u.l. [F/m]
d	Length [m]
f	Frequency [Hz]
f_n	Folding (Nyqist) frequency
G	Conductance [S]; also conductance p.u.l. [S/m]
g_i	Modal transfer functions
h	Time step
$I(s)$	Current in the frequency domain
$i(t)$	Current in the time domain
$\boldsymbol{I}_R$ and $\boldsymbol{V}_R$	Vectors of boundary currents and voltages at receiving end
$\boldsymbol{I}_S$ and $\boldsymbol{V}_S$	Vectors of boundary currents and voltages at sending end
$\boldsymbol{I}_{V_0} f(x)$	Interpolant in V_0
$\boldsymbol{I}_{W_j} f(x)$	Interpolant in W_j, $j \geq 0$
J	Wavelet level
k_i	i^{th} residue
l	Length [m]
L	Inductance [H]; also inductance p.u.l. [H/m]
L	Length of wavelet interval
l_K	Length of section K
L_n	Laguerre polynomials
m_i	i^{th} moment
N	Number of segments; total number of collocation points
$n_j = 2^j L$	Number of wavelet coefficients on j^{th} level

LIST OF SYMBOLS (continued)

$\boldsymbol{P}$	Transformation matrix
p_i	i^{th} pole
$\mathcal{P}_j f(x)$	Wavelet interpolant
$\boldsymbol{Q}$	Distribution matrix
q	Order of reduced model
R	Resistance [Ω]; also resistance p.u.l. [Ω/m]
s	Laplace variable
s_{ij}	Scattering parameters
T	Time period
t	Time
t_r	Rise time
$v(t)$	Voltage in the time domain
$V(s)$	Voltage in the frequency domain
Y	Admitance
$\hat{x}(t_1)$	Unknown wavelet coefficients (function of t_1 only)
Z	Impedance
Z_0	Characteristics impedance
Z_L	Load impedance

CHAPTER 1

Introduction and Problem Formulation

1.1. Introduction

In today's modern world, where speed is of the essence, the consumer is in constant pursuit of portable analog/digital electronics that are cheap, reliable and ultra-fast. There is no room for error or delay. To satisfy the consumer needs, a high level of integration at all levels of design hierarchy is required. This results in utilisation of deep-micron and multilayer packaging technologies. For example, current leading-edge logic processors have six to seven levels of high-density interconnect, and current leading-edge memory has three levels [ITRS99a]. Very Large Scale Integrated (VLSI) circuit complexity has already exceeded the 100 million transistors per chip and is continuing to grow [RC01].

Shrinking device features reduce the overall cost of the fabrication of an integrated circuit (IC) and at the same time enable operation at higher frequencies. A 180nm silicon technology with clock frequencies up to 720MHz is currently being replaced by a 90nm technology enabling clock frequencies up to 1.3 GHz. IBM, Intel and Texas instruments have presented their 65nm platforms and Freescale Semiconductor, Philips and STMicroelectronics have gone a step further by describing a 45nm technology [L04]. It is predicted that by 2011, a sub-50 nm technology will make it possible to have circuits operating at frequencies up to 2 GHz [DAR02]. The ever-increasing frequency blurs the once-distinct border between analog and digital design. It is predicted [DAR02] that in the future, no distinction between a time and frequency response will exist, i.e. digital, analogue and RF design will grow together.

When Moore [M65] observed an exponential growth in the number of transistors per integrated circuit and predicted that this trend would continue, very few scientists and engineers believed that the so called "Moore's Law", would hold true for long. But the main point of Moore's Law, the doubling of the number of transistors on a chip every couple of years, has been maintained until today. Naturally, the accompanying computer-aided design (CAD) tools need to improve at the same pace so that this progress can be sustained. However, the electronics industry worldwide faces increasingly difficult challenges today as it moves towards terahertz frequencies of

operation and with feature sizes in the nanometre scale. As the operating frequency grows by a factor of 5 every three years [D04], the previously negligible interconnect effects such as propagation delay, rise time degradation, signal reflection and ringing, crosstalk and current distribution related effects are now the principal issues for a circuit designer. If neglected during the design process, these effects can cause logic faults that result in the malfunction of the fabricated digital circuit. Alternatively, they can distort signals in such a manner that the circuit fails to meet its specifications [NA02]. Therefore, Electronic Design Automation (EDA) tools are employed in the early stages of design in order to take these high-frequency interconnect effects into account and avoid unnecessary and costly repeats of the design cycle [DAR03a], [DAR03b]. Some 60% to 70% of development time is currently allocated to simulation of a designed circuit [DAR03b] and it represents a major portion of the cost of a new product. The current trend of shrinking feature sizes and the increasing clock frequencies is expected to continue and it is envisaged that these signal integrity problems will continue to grow in the future. Hence, the development of adequate EDA tools that can, in an accurate and timely manner, address existing and emerging signal integrity issues is a prerequisite for electronic industry growth. Today, the design of accurate and efficient EDA tools is a critical research area.

1.2. Challenges facing the EDA community

The developers of circuit analysis algorithms are facing various challenges [D04] that have to be addressed in order to meet the demand of IC designers today. The *frequency challenge* relates to the wave character of signal propagation at ultra-high frequencies; thus an **accurate and efficient modelling of interconnect** is of paramount importance for successfully addressing the signal integrity issue in modern circuit design. The *functionality challenge* tackles the mixed analog/digital simulation issue. Very often a high-speed digital clock drives a relatively slow analog part of an IC. Specialised **envelope transient analysis methods** are necessary to yield acceptable results within a reasonable amount of computational time. The *shrinkage challenge* is concerned with the lack of a compact modelling approach as the feature sizes reach nanometre scale. Associated with the shrinkage problem is the *power challenge*. The reduction in feature size and the lower voltage levels of the power supply lead to a rising power density and a reduction in the signal-to-noise ratio thus necessitating

computationally expensive noise analysis. The EDA community needs to address these issues in order to ensure reliable and efficient design of new electronic products.

1.2.1. Frequency challenge

With an ever-increasing need for the timely arrival of information (e.g. in data transfer applications) there is a constant requirement for higher and higher operating frequencies. Every three years, the operating frequency of a chip increases by a factor of five and at the moment, typical rise/fall times and gate delays of an IC are under 50 ps [D04]. The frequency of a voltage-controlled oscillator (VCO) has already reached 50 GHz with the trend suggesting further increases. At these frequencies, the wave character of signal propagation becomes important and the signal integrity issue is the most important issue for the IC designers today.

It is out of the question to assume "ideal" connections between circuit elements today. Simple RC and RLC approximations just do not work at nowadays high frequencies. Designers have to treat interconnects as distributed networks, i.e. as transmission lines. A quasi-TEM mode of electrical signal propagation through an interconnect is assumed. The behaviour of interconnect is then described via the partial differential equations known as the Telegrapher's Equations that involve (in general) frequency-dependant per-unit-length parameters. Additionally, interconnect structures of the modern IC are non-uniform lines due to the complex geometries involved. Intensive computational efforts are necessary for simulation of circuits incorporating non-uniform transmission lines with frequency-dependant parameters.

Hence *there is a need for an efficient and accurate modelling strategy for non-uniform interconnect networks with frequency-dependant parameters.* This issue is addressed in Chapter 4 of this thesis and a novel method for simulating such interconnects based on a resonant analysis model of transmission lines is presented. Additionally, a method for efficient modelling of such interconnects characterised by a set of tabulated data is proposed in Chapter 5.

1.2.2. Functionality challenge

Modern ICs are becoming more and more complex with the latest trend being a complete system on a single chip (SoC). For example, a chip for a mobile phone may have an analog part (e.g. transmitter, receiver, etc.), a digital part (signal processing) and a memory (e.g. for phone address book) all in one chip. With such a complex

mixture of profoundly different parts, the problem of mixed analog/digital simulation arises. The ever-growing demand of the electronic industry for faster and smaller structures puts enormous demands on the numerical efficiency of such simulations. Today's focus is on using various *multi-time (multi-rate) schemes* to exploit latency in the different building blocks and hence speed up the simulation process. Other important functionality issues are *verification* of the analog part and *diagnosis* in case of failure.

Multi-time schemes. In mixed analog/digital circuits, a high-speed digital clock drives a relatively slow analog part of the IC. Therefore a long and very time-consuming *transient analysis* is necessary in order to capture both the high-frequency behaviour of the digital part and the low-frequency behaviour of the analog part. *This multi-scale problem requires specialised methods*, e.g. an envelope solver or a multi-time scheme, *in order to perform the simulation within acceptable time constraints*. This issue will be addressed in Chapter 8 of this thesis where a novel wavelet-based method for envelope simulation of non-linear circuits is proposed. In addition, this new envelope solver is extended so that it has the potential to greatly reduce the overall design cycle. This is possible due to the internal structure of the method that enables reuse of previously calculated results to obtain a more accurate transient response as explained in Chapter 9.

Verification and diagnosis. For digital modelling, formal verification is a well-established and much-needed area. However, verification procedures for analog modelling are very rare and insufficient. This is mainly due to the fact that both inputs and outputs are continuous. Thus much more effort is needed in the development of analog verification procedures, especially for high frequency applications.

If a simulation of a large circuit fails to converge, it is up to the designer to identify the flaw in the circuit design and correct it. Thus, the simulation algorithm has to provide relevant information about the conditions under which the simulation failed so that the designer can rectify his design.

Although both verification and diagnosis are important issues for the EDA industry, both of them are beyond the scope of this dissertation and will not be discussed any further.

1.2.3. Shrinkage challenge

The physical size of electronic circuits is rapidly shrinking. From 700nm technology in 1990, manufacturing technology had reduced to 350nm in 1995. The year 2000 has seen the introduction of 180 nm technology and 90 nm is a reality these days (2005), with the 65 and 45 nm technology just around the corner [L04]. However, a reduction in physical size has brought new problems. The shrinking size of the device requires *more physical effects to be included* into a model and hence the model complexity becomes such that simulation times and storage requirements became impractical.

Electromagnetic device modelling. Inclusion of more physical effects into a model means that previously negligible effects of the electrical and magnetic field may be required to be taken into account. The standard circuit description of a device in terms of port currents and voltages does not provide a framework to accurately describe device behaviour at high frequencies when the influence of electromagnetic fields becomes a substantial factor in the overall device response. In such cases a device has to be described in terms of Maxwell's Equations. However, this necessitates a computational effort that is significantly greater than for circuit modelling. An additional concern is the definition of a criterion for selecting the appropriate model to be employed, that is whether to describe the device in terms of currents and voltages or in terms of electrical and magnetic fields.

Coupling between device and circuit simulation. As discussed, full electromagnetic device simulation may be needed for some critical circuit components in the modern IC. In this case a set of partial differential equations, i.e. Maxwell's Equations, govern device simulation. On the other hand, a set of ordinary differential equations governs circuit simulation. Therefore, it is necessary to combine these two types of differential equations in order to obtain an overall simulation result. However, obtaining one common solution to a mixture of two distinct types of differential equations is a complex problem that requires a carefully designed numerical approach [SM03].

It should be noted that, although significant, electromagnetic device modelling and the coupling between device and circuit simulation are beyond the scope of the current contribution and will not be investigated any further.

1.2.4. Power challenge

The shrinking in the size of the devices results in an *increase of the power density* since the switching currents are confined within smaller areas. This makes the chip more susceptible to thermal failure. The problem introduced by the increase in the power density is partly compensated by the recent trend of a *reduction in the power supply voltage* from 5.0V to 3.3V and further down to 1.0V and lower. Furthermore, lower power supply voltages enable ICs to operate at even higher frequencies. However, the decrease in the power supply voltage level has also led to a *decrease in the signal-to-noise ratio*, which in turn means that *parasitic effects, noise influence* and *power leakage* on the overall chip performance have increased. For example, in 90nm technology power leakage accounts for almost 50% of chip power consumption [E04]. In addition, reduction of the supply voltage increases crosstalk problems.

Power density problem. The shrinkage in the feature size and the reduced power supply level result in an increase in the power density which can be roughly approximated as [D04]:

$$power\ density \sim \frac{power\ supply}{(shrink\ factor)^2}$$

As can be seen, the effect of the increase in the power density due to shrinking size is partially compensated by the decrease in the power supply. At the moment, power density is above 100 Watt/cm^2 [D04]. Increasing power density results in two major problems: *how to cool the chip* and the problem of the so called *"hot spots"*, the parts of the chip that are too hot while the average temperature is still within specified limits.

From the simulation point of view, intelligent coupling between circuit and thermal simulation is necessary. Using a direct simulation approach yields very long transient simulations even for small circuits. For large circuits, this computational effort is very large and the simulation time may be unacceptably long. This is due to the fact that the time constants of the thermal process and the circuit operation differ by 3 to 6 orders of magnitude. Hence, there is a need for a multi-rate method that will enable intelligent coupling between thermal and circuit simulation. Although the thermal problem has not been investigated in this thesis, a multi-time wavelet-based envelope solver proposed in the Chapter 8 might be used in this context as well.

Parasitic effects and noise analysis. The reduction in the power supply voltage level and the shrinking of the physical size of ICs has lead to a reduction in the signal-to-noise ratio for modern chips, thus making them more susceptible to noise and the influence of parasitic effects. The need for a better description of parasitic effects necessitates a greater level of accuracy in the parasitic extraction process. Use of additional resistances, capacitances and inductances in the model has led to a significant rise in the number of nodes and increases in the dimension of the system matrices. Because the fill-in sparsity in the system matrix is decreased, the simulation effort due to the increased number of linear algebra calculations is increased. A reduction in the computational complexity of simulations that include these parasitic effects is the subject of ongoing research efforts.

The issues that the EDA industry is required to address are diverse and complex. The current trends of ever-rising operational frequencies and shrinking feature sizes result in two major requirements for simulation tools: maintaining high *accuracy* while making sure that the *efficiency* of the numerical calculations is acceptable. Inevitably, trade-offs need to be made. This thesis addresses the *frequency and functionality challenge*. The related issues of shrinkage and power challenges are beyond the scope of the research presented here but nevertheless their importance should not be disregarded.

1.3. Existing simulators

To simulate a complex electronic circuit, a suitable computer aided design (CAD) simulator is employed. The existing CAD simulators may be classified into two groups: electromagnetic (full-wave) simulators and circuit solvers.

1.3.1. Electromagnetic (full-wave) simulators

With the increase in the operating frequency the field effects can become substantial and cannot be neglected [RC01]. Thus, when a full accuracy is required, an *electromagnetic simulator* that solves Maxwell's Equations [P94] is used. In this case, the system behaviour is described in terms of time- and space-dependant values of electric field intensity (***E***), magnetic field intensity (***H***), electric flux density (***D***), magnetic flux density (***B***) and distributed current sources (***J***). Since Maxwell's theory is general (i.e. does not neglect field effects), electromagnetic simulators provide better simulation accuracy than standard circuit solvers. The price to be paid is in terms of

increased computational complexity and often-unacceptably long simulation times (from a couple of hours to a few days). Hence, full-wave simulators are not fast enough to be used in the everyday design tasks. Full wave simulators such as Ansoft HFSS, Cosmos HFS 3D, Quickwave 3D, etc. are employed only when full accuracy is absolutely necessary.

1.3.2. Circuit simulators

Circuit simulators use modified nodal analysis (MNA) matrices [HRB75] that describe a system based on Kirchoff's Theory. The system behaviour is described in terms of time-dependant (but not space-dependant since the field effects are assumed to be negligible) values of currents (I) and voltages (V) and the topology of a circuit is given via a lumped element representation (resistors (R), capacitors (C), inductances (L) and admittances (G)). Distributed systems (e.g. interconnects that behave as transmission lines at high frequencies) may be taken into account through derived "stamps" for inclusion in the appropriate matrix [AN01]. Circuit simulators are capable of very efficient simulation of very complex circuits typically requiring from a few seconds to a few hours to obtain a result. However, at today's high frequencies, new demands are being placed on existing circuit simulators.

Not long after the introduction of the first commercial IC in 1961 (Fairchild and Texas Instruments), it was recognized that the computer would play a central role in the design and analysis of integrated electronics. It started in 1967 when Bill Howard made the first implementation of a computer program (BIAS) for the analysis of the nonlinear dc operating point of an IC [N95]. The milestone in the circuit simulation industry was the development of CANCER (Computer Analysis of Nonlinear Circuits Excluding Radiation) [NR71] in 1971. This result of a class project at Berkley was a starting point for the first truly public-domain, general-purpose circuit simulator called SPICE (Simulation Program with Integrated Circuit Emphasis) which was released in May 1972. SPICE continued to improve and SPICE2 became a reality in 1975. The latest version of SPICE (SPICE3), written in the C programming language instead of FORTRAN, was released in 1985. SPICE from Berkley has been freely available and many argue that this fact, along with the quality of software, is the key factor in its worldwide popularity. SPICE is the godfather of many current commercially available simulators such as HSPICE (from Avant!), PSPICE and Spectre (Cadence, formerly Orcad), APLAC (APLAC Solutions Inc.) and HISM (Nassda Corporation), as well as

in-house developments TITAN (Infineon), TI-SPICE (Texas Instruments), AS/X (IBM) and P-STAR (Philips).

Until recently, the success of SPICE was unmatched. SPICE simulations were universally applicable and yielded realistic and reliable results. But the complexity of a typical integrated circuit has grown enormously. As the size of a single device in the IC is getting smaller, the number of the devices in a single chip is growing. Smaller devices necessitate ever more complex device models; the large number of devices makes the time necessary to perform the overall simulation unacceptably long. Observing current trends in circuit modelling Nagel, one of the pioneers of SPICE, asks "*Is it time for SPICE 4*"? [N04]. The amount of research efforts into overcoming the current challenges in circuit simulation implies that the answer is most definitely *yes, there is a need for 21st century circuit simulator.*

It should be noted that the research efforts in this thesis are restricted to advances in the state-of-art in circuit simulators and from this point on, only issues related to circuit simulators will be discussed.

1.4. Thesis objective and contributions

In order to address the problem of accurate and efficient transient simulation of a complex electronic circuit, the standard approach is to identify two integral parts: a nonlinear network $\mathcal{N}$ and a linear interconnect network $\mathcal{L}$ as presented in Fig 1.1.

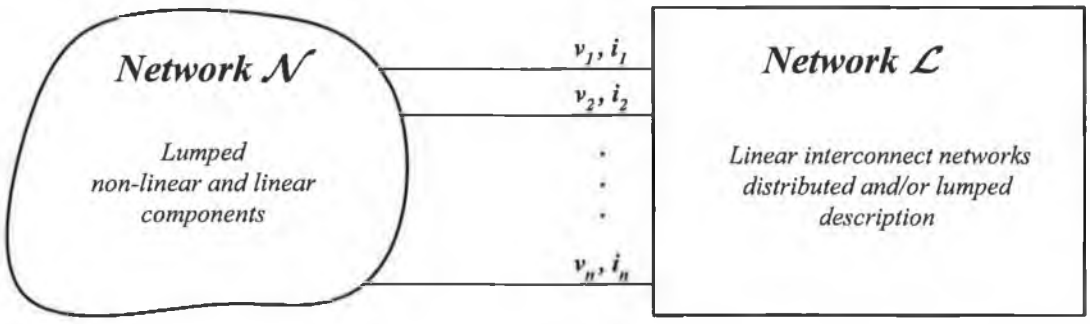


Fig 1.1. *A high-speed complex electronic system*

The specific issues associated with their simulation may then be addressed separately taking into account the nature of the elements involved. Chapters 2 to 5 are concerned with the issues arising from simulation of linear interconnect networks. Chapters 6 to 9 address the issues arising from simulation of non-linear circuit elements.

Specifically, in Chapter 6, numerical algorithms for obtaining the solution to a set of stiff ordinary differential equations that describe the behaviour of high-frequency non-linear circuits are discussed.

1.4.1. Research objective

The main objective of the research that is presented in this thesis is to advance the state-of-art in *transient simulation* of *complex electronic circuits* operating at *ultra high frequencies*. Given a set of excitations and initial conditions, the research problem involves determining the transient response of a high-frequency complex electronic system consisting of a linear and non-linear part:

- with greatly **improved efficiency** compared to existing methods
- with the potential for very **high accuracy**
- in a way which permits a cost-effective **trade-off** between accuracy and computational complexity.

The proposed advances are summarised in the following section.

1.4.2. Thesis contributions

This section summarises the proposed contributions of this dissertation. They have been categorised under three headings: linear subnetwork simulation (L), numerical algorithms for the transient analysis of high frequency circuits (A) and non-linear circuit simulation (N).

1.4.2.1. Linear subnetwork simulation

Modelling of complex *linear* interconnect networks has received a lot of attention recently due to the need to properly capture the frequency-dependent behaviour of interconnect structures operating at high-frequencies.

The approach proposed in this dissertation is based on a transmission line (TL) model centred around *natural modes of oscillation* of a line [WC97]. Initially, the resonant model that describes the transmission line is formed in the frequency domain thus enabling the capture of frequency-dependent parameters. As described in Chapter 4, the particular model construction procedure is such that it does not require the assumption of uniformity of the transmission lines, hence non-uniform interconnects can readily be described with this model. This resonant model has two distinct advantages: 1) it enables a straightforward transfer of the frequency-domain model to its

time-domain counterpart with a minimal loss of accuracy; 2) the internal structure of the resonant model is such that the efficiency of numerical calculations may be greatly improved using a suitable model order reduction technique.

The following are the contributions regarding *linear subnetwork simulation* that are presented in this thesis:

L1) A model order reduction technique for the resonant model based on neglecting higher modes of oscillation on the transmission line is presented. A detailed description and reasoning behind it is described in detail in Section 4.3. Transient responses from a full and reduced model are obtained and compared. Excellent agreement between the transient response of a full model and reduced model will be shown. The error distributions are presented and the model bandwidth is disscussed.

L2) A very efficient technique for interconnect simulation is presented in Section 4.4. It combines in an original manner a model order reduction technique based on the Lanczos process [AS00] with the resonant model. Transient responses for two illustrative examples, a single interconnect system with frequency-dependant parameters and a coupled interconnect system, have been obtained for both a full-sized and reduced-sized system. As evidenced by results published in [CD03] and [DC03], significant gains in terms of computational time and memory resources have been achieved without compromising the accuracy of the output.

L3) It is not always possible to derive analytical models for interconnects due to the complexity and the inhomogeneity of the geometries involved. In such cases, the interconnect networks are usually characterised by frequency-domain parameters derived from measurements or rigorous full-wave simulation. The novel method proposed in Chapter 5 of this thesis and published in [CDB05] is capable of generating highly accurate macromodels in the time domain from the available measured or simulated frequency-domain data. Therefore, the method proposed is independent of the interconnect geometries involved. The efficiency of the method is further improved by utilizing a judiciously chosen Laguerre model order technique [CBK+02].

1.4.2.2. Numerical algorithms for the transient analysis of high frequency circuits

The simulation of a high frequency non-linear system requires at some point that a numerical solution to a system of typically highly non-linear differential equations is found. Usually these equations arise from non-linear equivalent circuit models for microwave active devices. The character of the device equivalent circuit models is such that 'stiff' ordinary differential equations are often found due to the widely varying time constants in the non-linear circuit. The short time constants force the simulator to operate at an extremely small calculation step for the entire time scope of the simulation although the influence of these elements usually becomes negligible after few simulator steps. This seriously hinders the efficiency of the simulator in general. Thus there is a need for new numerical methods specially designed for solving *stiff* ODEs that take into account the nature of elements involved.

In total, **four new** methods for obtaining the solution to stiff ODEs are developed and presented in Chapter 6 of this thesis. The basic idea behind these methods is similar to that of [GN97], where a sequence of local Padé approximations to the solution of the ODE is built in order to provide a solution to the ODE. The method is then advanced in time by using the solution at a specific time point as the initial condition for the next time-step.

The following are the contributions relating to *numerical algorithms for solving stiff ODEs* that are presented in this thesis:

A1) Proposed Exact-fit and Padé-fit methods are multistep methods that do not require obtaining higher order derivatives of the function describing the ODE. It is recommended to use them in cases where the analytic expression for the function is very complicated. Additionally, the corrector formulas for use in a predictor-corrector setup are derived.

A2) Padé-Taylor and Padé-Xin are singlestep methods that require obtaining higher order derivatives of the function describing the ODE. The Padé-Taylor corrector formula for use in a predictor-corrector setup is developed and numerical results are published in [CDB02].

1.4.2.3. Non-linear circuit simulation

Very often high-speed digital signals drive relatively slow non-linear analog parts of an IC. This results in long simulation times to capture a complete response. Frequently, the complexity of the designed electronic circuit is such that it is simply not possible to perform such analysis using standard techniques within the time allocated for the design of a new circuit. Therefore, specialised methods for transient analysis of circuits that have parts with widely-separated time constants are necessary.

The following are the contributions regarding *non-linear circuit simulation* that are presented in this thesis:

N1) A novel approach for the simulation of high-frequency circuits carrying modulated signals is developed and presented in Chapter 8. The approach combines a wavelet-based collocation technique with a multi-time approach to result in a novel simulation technique that enables the desired trade-off between the required accuracy and computational efficiency. This work is published in [CD03b].

N2) To further improve the computational efficiency of the wavelet-based approach, a non-linear model-order reduction (MOR) technique [GN99] is applied to the approach in N1). This results in a highly efficient circuit simulation technique specially suited for highly nonlinear circuits with widely-separated time constants as presented in Section 8.5. Furthermore, a trade-off between the desired efficiency and required accuracy is easily achieved by simply adjusting the wavelet level depth and reduction factor as evident from the results published in [DCB04a].

N3) Based on the approach N2), a novel wavelet-based method for the analysis and simulation of IC circuits with the potential to greatly shorten the IC design cycle is developed and presented in Chapter 9. The preliminary phase of a design process involves obtaining an initial result for the circuit response to verify the functionality of the design. For this purpose, the previously presented wavelet-based approach N2) is utilised. Then, when a higher degree of accuracy is sought for fine-tuning of the designed IC, the previously obtained numerical results are then reused to compute the more detailed transient response results as reported in [DCB05]. The major saving in the design time is obtained by avoiding a restart of the complete simulation from the beginning. Instead, based on the coefficients obtained from an

initial calculation, only the coefficients necessary for the next level of model accuracy are computed. This results in a substantial shortening of the overall design cycle.

N4) The efficiency of method in N3) is further improved by using the same non-linear model order reduction technique in the process for obtaining the more detailed results as presented in Section 9.3 and published in [DCB04b].

1.5. Thesis overview

This thesis presents advances in the *transient simulation* of *complex electronic circuits* operating at *ultra-high frequencies*. Given a complex electronic circuit to be simulated, specific issues associated with the simulation of a linear interconnects and general non-linear circuits are addressed and the results are reported in this dissertation.

The research contents and contributions are specified in Chapter 1.

In Chapter 2 some basic background regarding interconnects is introduced. A short description of interconnect effects and their influence on the integrity of high-speed signals propagating through an interconnect is presented. Some available interconnect models are described and important simulation and mathematical issues are underlined.

The existing techniques for modelling and simulation of high-speed interconnects may be roughly classified into two groups: strategies based on transmission line macromodelling and interconnect modelling techniques based on model order reduction approaches. The basic principles and advantages/disadvantages of these techniques are given in Chapter 3.

Chapter 4 is concerned with the development of interconnect models from a Telegrapher's Equations description. Initially, a resonant model in the frequency domain is formed thus capturing frequency-dependant characteristics of either uniform or non-uniform interconnect. After conversion to the time domain, a model order reduction technique is applied resulting in two highly efficient interconnect simulation techniques. Experimental results that are presented here confirm both the accuracy and the efficiency of the proposed approach. Related publications: [CD03a] and [DC03].

However, an interconnect description may not always be available in analytical form due to its complex structure and geometry. In such cases, the interconnect networks are usually characterised by a set of tabulated data. The data is usually in the form of frequency-domain scattering parameters derived from measurements or rigorous full-wave simulation. A novel method for the simulation of interconnects described via a tabulated data set is presented in Chapter 5. Experimental results obtained for two sample circuits validate the approach. Related publication: [CDB05].

Results from investigation into numerical algorithms for the transient simulation of high-speed circuits are presented in Chapter 6. In total, four new methods for solving stiff ODEs are developed. Related publication: [CDB02].

An introduction to the area of wavelets is provided for the reader in Chapter 7. Some basic notations are introduced and a brief discussion on some wavelet-related issues is given. Finally, a wavelet-like basis that is used for development of a novel envelope transient analysis technique is given.

In Chapter 8, a novel wavelet-based approach for envelope simulation of circuits carrying signals with widely separated time scales is presented. This approach combines a wavelet-based collocation technique with a multi-time approach to result in a novel non-linear circuit simulation technique. A non-linear model order reduction (MOR) technique is applied to speed up the computations. The main advantage of the proposed technique is that it enables the desired trade-off between the required accuracy and computational efficiency. Related publications: [CD03b] and [DCB04a].

A simulation technique that enables a reduction in the design cycle time is presented in Chapter 9. Initially, the transient response is obtained with the method described in Chapter 8 so that the correct functionality of the designed circuit may be verified. Later on, when a higher degree of accuracy for fine-tuning the designed IC is sought, the initial numerical results are reused for obtaining highly-accurate results. The method offers major savings in design time and ultimately enables avoiding costly time-to-market delays. Related publications: [DCB04b] and [DCB05].

Finally in Chapter 10, a summary of the research carried out for this thesis is presented. Suggestions for possible extensions and a discussion as to how this work might continue are given.

CHAPTER 2

Simulation of High-Frequency Integrated Circuits

As microprocessor clock speeds continue to rise above the gigahertz mark and the physical size of transistors is already expressed in nanometres [C04], interconnects are emerging as the major bottleneck in the growth of VLSI technology. The influence of electromagnetic and distributed effects of an interconnect on the overall performance of high-speed VLSI chips is the key difficulty that has to be addressed in a timely and accurate manner. Interconnect effects such as propagation delay, crosstalk and skin effect are proven to be the major cause of signal degradation in high–frequency circuits [DCK+01], [AN01], [D98], [G94], [JG93]. If not taken into account during the design stage of a high-frequency circuit, interconnect effects can cause serious misrepresentation of logic levels in a prototype of a designed digital circuit or they can deform the analog signal in such a manner as to render the fabricated circuit worthless [NA02]. Better than 10% accuracy in the prediction of signal distortion due to interconnect effects is necessary to ensure the correct operation of the designed IC [CCH+01]. As a result accurate modelling of interconnects becomes an essential part of a design process and interconnect analysis is a requirement for all state-of-art circuit simulators today.

This Chapter aims to review several background topics regarding the simulation of *high-frequency* (HF) integrated circuits. First, the term "*high frequency*" will be explained and subsequently, the term "*high-frequency interconnect*" in the framework of this thesis will be defined. An overview of *interconnect effects* and their effect on a signal propagating through HF interconnect will be given. A general review of existing *electrical models* for HF interconnect will be presented. Finally, some important *interconnect simulation issues* will be highlighted.

2.1. High-frequency interconnect

Prior to addressing the design problems of high speed interconnects, it is necessary to define what is an interconnect. The Penguin Dictionary of Electronics [PDE88] states that *interconnect* is:

- *Any method of providing an electrical path between any of the materials (metals, semiconductors, etc.) that combine to form a circuit.*
- *Connections between and external to any functional item that form a circuit or system of circuits. Functional items include component parts, devices, subassemblies and assemblies.*

The function of interconnects is to distribute clock and other signals and to provide power/ground to and among the various circuit/system functions on the chip [ITRS99a]. An interconnect can be found at chip level, printed circuit board (PCB), multi-chip modules (MCM), packaging structures and backplanes [AN01]. With such a variety of interconnect structures present today, it is an enormous challenge to develop a general interconnect simulation tool that can accurately and efficiently describe the behaviour of an arbitrary interconnect.

In early days of integrated circuit (IC) technology, designers were not concerned with the interconnections between the lumped elements that incorporated the main functionality of the designed chip. They simply chose to disregard any influence interconnects might have on a signal transmitted through them, thus, in effect, considering them as a short between the two circuit elements they were connecting. This assumption eased the design process and it seemed to be justified – the measured results did not show much discrepancy with the predicted ones. But the rising operational frequency and shrinking device size caused interconnect to gradually display effects that are responsible for degradation of a signal propagating through them. Thus these high-frequency interconnect effects have to be taken into account during a design process in order to ensure the high-quality of overall chip's performance.

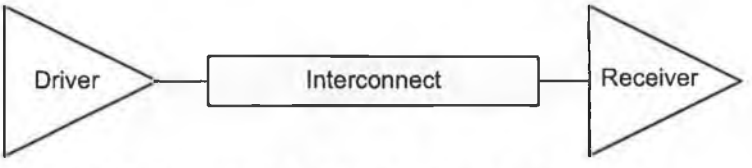


Fig. 2.1. *Interconnect in relation to driver and receiver circuit*

So what is *high-frequency interconnect*? The answer to this question can be observed either in the time- or the frequency-domain [AN01]. The speed of an electrical signal propagating through an interconnect is extremely fast but finite. Hence, it needs some time to propagate through an interconnect and the longer the interconnect is, the more time the signal needs to reach its end point. Once the signal's rise/fall time is

approximately the same level as its propagation time, interconnect may not be considered anymore as a short between the driver circuit and the receiver circuit [AN01], [B90], [JG93]. Instead, within the rise/fall time of signal, the impedance of interconnect becomes the load for the driver and also the input impedance to the receiver circuit as illustrated in Fig. 2.1. Achar and Nakhla [AN01] define the *high-frequency interconnect as the one in which the time taken by the propagating signal to travel between its end points cannot be neglected.*

High-frequency interconnect may also be observed in the frequency domain in terms of the frequency content of signal propagating through it [AN01]. At low frequencies, an interconnect behaves as an ordinary wire, that is, connecting two circuit components without any obvious change in the signal spectrum. But as the frequency of the propagating signal rises the resistive, capacitive and inductive properties of an interconnect come into play [DKR+97], [DCK+01]. Due to these, the frequency content of a signal is altered and signal may become distorted. In addition, faster clock speeds and sharper rise times are adding more and more high-frequency content to the spectra of the propagating signal. Thus it can be said that *a high frequency interconnect is one that considerably influences the frequency spectrum of a propagating signal.*

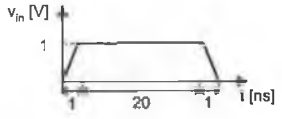
In summary, the key characteristics of a high-frequency interconnect is that it distorts the properties of a propagating signal both in the time and the frequency domain. Henceforth the effects that cause distortion of a signal propagating on high-frequency interconnect will be referred to as the high-frequency *interconnect effects.* Furthermore, Matick [M69] showed that any two uniform parallel conductors that are used to transmit electromagnetic energy could be considered as transmission lines. Hence all transmission line theory concepts are readily applicable to the analysis of high-frequency interconnect behaviour.

2.2. High-frequency interconnect effects

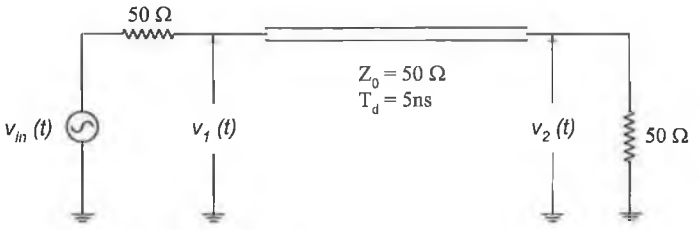
With the rapid advancement in IC technology, numerous interconnect effects such as propagation delay, attenuation, crosstalk, signal reflection, ringing and current distribution effects have become important factors during the design stage. Therefore, their inclusion in the simulation of a circuit is an absolute necessity in circuit design. This section presents an overview of interconnect effects and their influence on the shape of a propagating signal. The examples used to illustrate the effects are taken from [AN01], [NA02].

2.2.1. Propagation delay

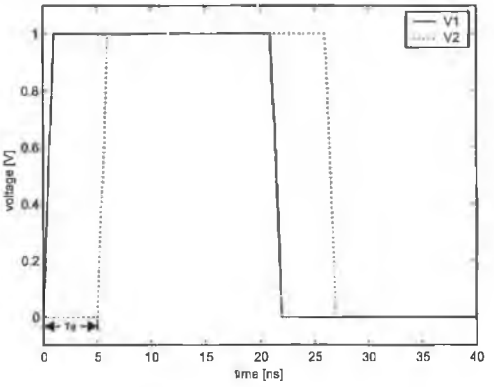
The effect of propagation delay is a direct consequence of the fact that a signal propagates through an interconnect in some finite time. If that time is much less than the time constants of the discrete circuit components that the interconnect is connecting, it can be considered that signal propagation was instantaneous and no distortion of the propagating signal occurred. However, if the time the signal takes to traverse through the interconnect is comparable with the time constants in the system, the propagation delay cannot be neglected as it may seriously influence the signal properties. Fig. 2.2 illustrates propagation delay in the case of a lossless interconnect that acts as an ideal delay line.



a) Input voltage



b) Network with lossless interconnect



c) Transient Response

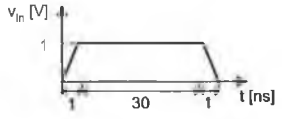
Fig. 2.2. *Illustration of propagation delay*

The propagation delay emerged as a serious problem for the first time in 250nm technology designs where the signal delay between the logic cells is heavily influenced by the capacitance and resistance properties of the wires connecting the logic gates [E04]. With the ever shrinking sizes of the manufacturing technology, the issue of signal delay has become particularly important. For example, in a 130 nm design, the interconnects are responsible for more than 75 % of overall delay on a chip [E04]. In order to predict signal delay, RC models of interconnect were initially used. However,

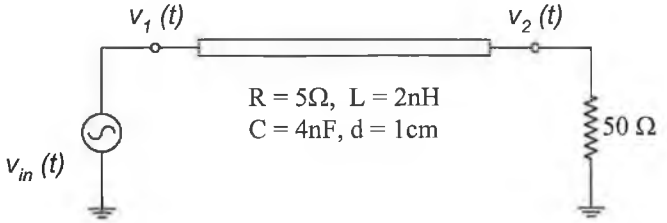
these models greatly overpredict signal delay resulting in the use of larger devices than necessary. These have higher power consumption and generate more crosstalk than otherwise would be the case [DCK+01]. Recently, the use of distributed RLC circuits for more accurate signal delay prediction has become the norm [DKR+97], [DCK+01].

2.2.2. Rise time degradation

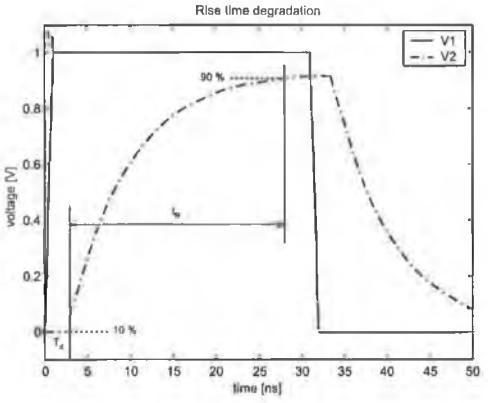
The current design trend of utilising short lines wherever it is possible has resulted in signal delay as less of an obstacle than it used to be. Today, the closely related problem of *rise-time degradation* has become the more important factor in obtaining even faster circuits [JG93], [W04]. In general, rise-time is defined as the time taken by the signal to rise from the 10% to the 90% of the final voltage level [NA02]. The rise time degradation occurs when the rise time at the receiver end (t_R) is greater than the rise time at the source end (t_r).



a) Input voltage



b) Network with lossy interconnect



c) Transient Response

Fig. 2.3. *Illustration of rise time degradation*

As can be seen from Fig. 2.3, the rise time degradation may greatly increase the overall delay of signal propagation through the line. Hence, the minimum and maximum attainable digital logic levels between the switching intervals are heavily influenced by this effect. It is predicted that in the future the attainable rise time will be

the major limiting factor in achieving ultra-fast operational frequencies. Very sophisticated distributed interconnect models are necessary to accurately describe the rise time degradation effect [DCK+01].

From a practical point of view in most digital applications, the desired highest operating frequency of interest f_{max}, is related to the rise/fall time (t_r) of the propagating signal. For most signal pulses, the energy spectrum is spread over an infinite frequency range but most of the signal energy is concentrated in the low-frequency region and the energy content decreases rapidly with frequency. Hence, ignoring the high-frequency components of the spectrum above a maximum frequency f_{max} will not seriously alter the overall signal shape and for all practical purposes, the width of the spectrum can be assumed to be finite [IJ02], [OSB99]. On the other hand, sharp pulses contain high frequency harmonics that need to be taken into account and hence, the signal rise time will contain the highest frequency component of interest. This frequency f_{max}, in essence, defines the *bandwidth*, i.e. frequency range of interest, for a given interconnect.

In order to determine approximately the highest frequency component of interest in a propagating digital signal, one may consider an interconnect as a simple low-pass filter, i.e. RC circuit [DCK+01]. Then f_{max} may be defined as the "upper 3dB frequency" of such a filter, i.e. the frequency at which the gain falls off to 0.707 of its low frequency value, given by:

$$f_{max} = \frac{1}{2\pi RC}. \tag{2.1}$$

Since the rise time for such a representation with a step input [DKC+01] is $t_r = 2.2RC$, the following relationship results:

$$f_{max} \approx \frac{0.35}{t_r} \tag{2.2}$$

This relationship has been proposed for practical use by several authors [JG93], [DZ92], [KGP94], [D98], [AN01], although for some implementations [CPP+99] a more strict relationship (2.3) is suggested:

$$f_{max} \approx \frac{0.1}{t_r} \tag{2.3}$$

In order to avoid signal degradation, the pulse width (p_w) should be not less than the inverse of this f_{max} [D98], i.e.

$$p_w \geq \frac{1}{f_{max}} \quad \Rightarrow \quad t_r \leq 0.35 p_w \tag{2.4}$$

If this condition is not maintained, the pulse would have a sinusoidal-like shape and a large delay would exist in the system. In the case of a processor, the pulsewidth is equal to processor cycle time. For example, a digital system with a processor operating at 1GHz (processor cycle time is 1ns) requires a rise time of $t_r < 0.35$ns [DKC+01]. Deutch et al. [D98], [DKC+01] suggest that, due to crosstalk, reflections caused by discontinuities result in waveforms with the narrow peaks. These have higher frequency components than f_{max}, and hence a more practical frequency of interest might actually be 5 x f_{max}.

2.2.3. Attenuation

The signal propagating through interconnect is subjected to certain losses which can be either resistive or conductive in nature. Resistive losses are caused by today's design trend of a reduction in cross-section area. This increases the resistance of a line. Conductive losses are also a function of frequency and they are proportional to the dielectric loss factor of the dielectric material.

Both resistive and conductive losses have a very significant influence on propagation of a digital signal since they directly influence the logic levels of digital signals. If the level of attenuation is too high, the receiver circuit may fail to recognize the digital signal correctly and hence false switching can occur. Fig 2.4 illustrates the attenuation experienced by a lossy line in Fig. 2.3.b).

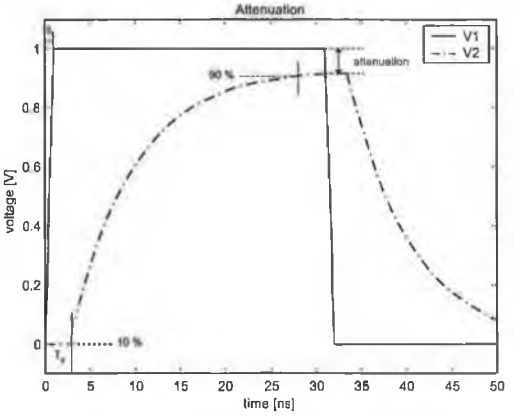


Fig. 2.4. *Illustration of attenuation*

As a guideline, Deutsch et al. [D98] recommend that the attenuation of the highest frequency component that has significant energy in the signal rise time should not exceed 3-5 dB. If this condition is not upheld, false switching in a digital IC may occur.

2.2.4. Reflection and ringing

Signal reflection and the associated ringing are other high frequency interconnect effects that can cause severe degradation of a propagating signal. The discontinuity in the characteristic impedance of the transmitting line and the impedance mismatch between the line characteristic impedance and source/terminating impedances are two major causes of signal reflection and ringing [AN01].

Some common causes of a *discontinuity in characteristic impedance* include connectors between card-to-board, cable-to-card, leads between chip and chip carriers, or between card wiring and chip carriers, long vias, orthogonal wiring, wire bonds and redistribution lines [D98], [AN01]. If the delay on the discontinuity is much smaller than the signal rise time t_r, the degradation of the signal is negligible. However, if the delay is close to half of the rise time, the waveform experiences a significant degradation due to the reflections.

The *impedance mismatch* between the line characteristic impedance and source/terminating impedances causes effects such as undershooting, overshooting and ringing. Consider the simple case of the impedance mismatch shown in Fig. 2.5.

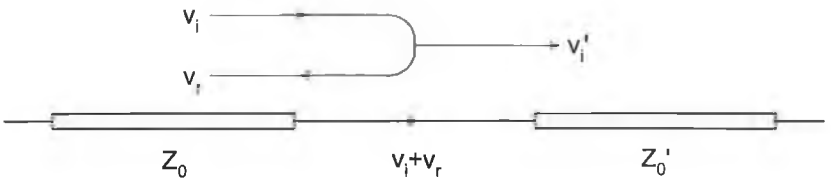
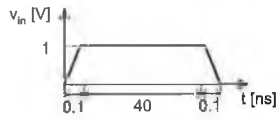


Fig. 2.5. *Impedance mismatch*

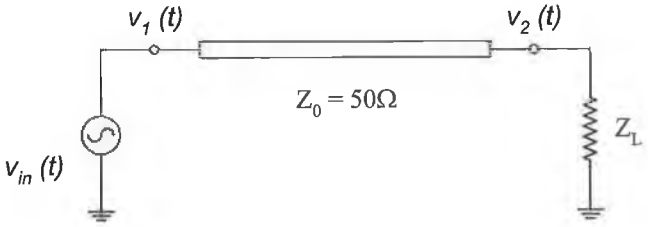
The impedance variation from Z_0 to Z_0' causes part of the onward propagating signal v_i to be reflected (v_r) with the reflection coefficient (ρ) given as [P94]:

$$\rho = \frac{v_r}{v_i} = \frac{Z_0' - Z_0}{Z_0' + Z_0} \tag{2.5}$$

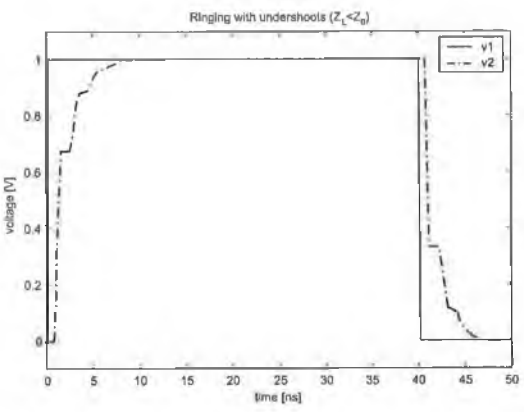
If $Z_0' = Z_0$, i.e. the impedances are matched, then the reflection coefficient is equal to zero ($\rho = 0$). This means that $v_r = 0$, i.e. no reflection occurs. However, when impedances are mismatched ($Z_0' \neq Z_0$), the reflection coefficient $\rho \neq 0$ and part of the electrical wave is reflected back causing under/overshooting shown in Fig. 2.6 and ringing effects as shown in Fig 2.7.



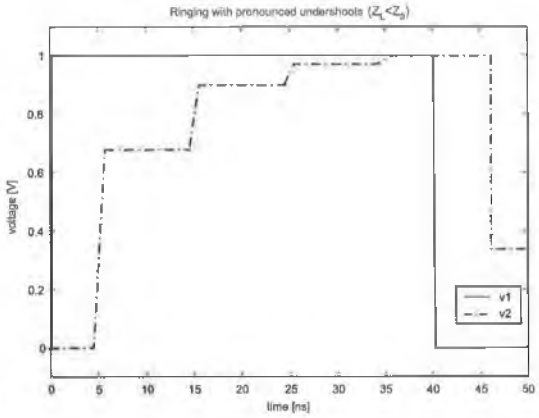
a) Input voltage



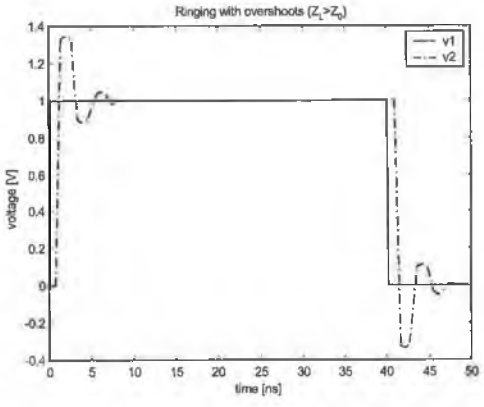
b) Network with lossless interconnect



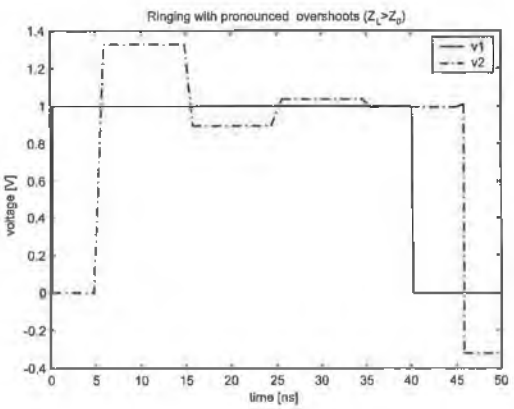
c) T_d=1ns, Z_L=25 Ω



d) T_d=5ns, Z_L=25 Ω



e) T_d=1ns, Z_L=100 Ω



f) T_d=5ns, Z_L=100 Ω

Fig. 2.6. *Illustration of undershoots and overshoots in lossless interconnect*

As seen from Fig. 2.6 if the terminating impedance Z_L is smaller/higher than the characteristic impedance Z_0 of the interconnect undershooting/ overshooting occurs.

Fig. 2.7 shows the effect of ringing in the lossy line for different cases of terminating impedance. It is clear that the ringing effect becomes more pronounced for higher values of terminating impedance Z_L.

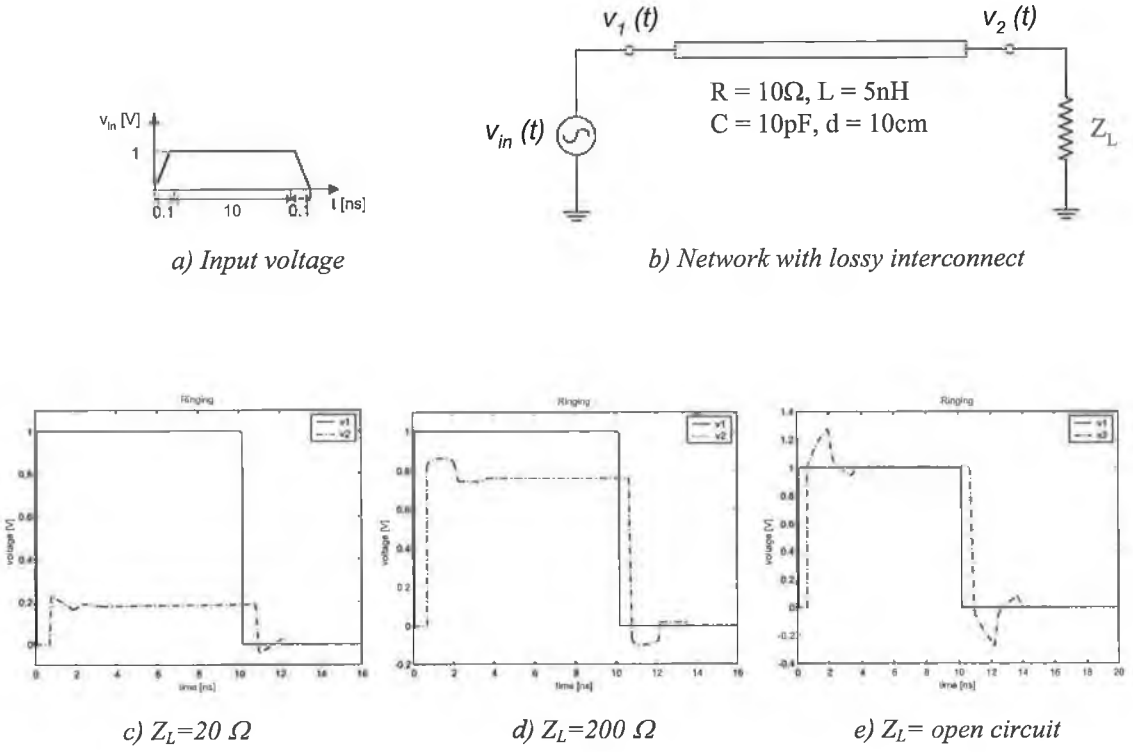


Fig. 2.7. *Illustration of ringing in lossy interconnect for various cases of termination*

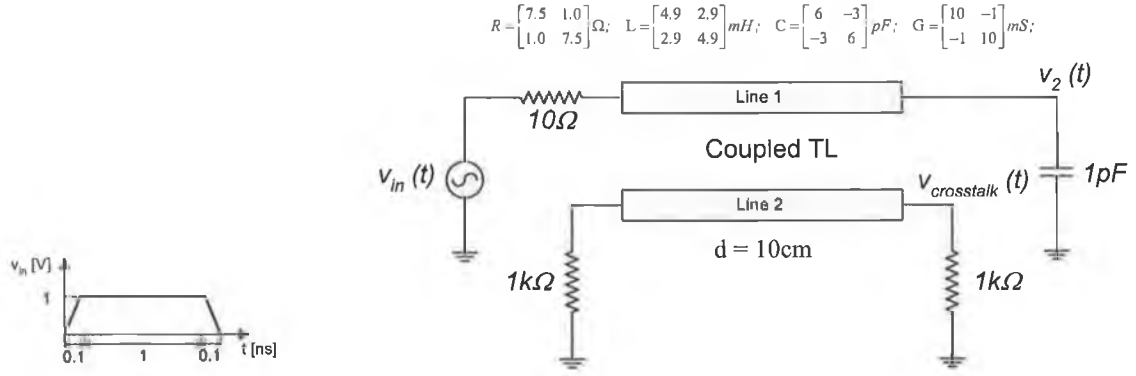
As evidenced from Fig 2.6 and Fig 2.7 signal reflections and ringing can severely distort the propagating signal. A designer needs to take into account these effects when designing a new IC. Thus an interconnect model must be able to simulate these effects correctly. Since these effects are non-monotonic in nature, as a minimum requirement the model has to be able to describe non-monotonic behaviour, i.e. simple RC approximations are insufficient.

2.2.5. Crosstalk

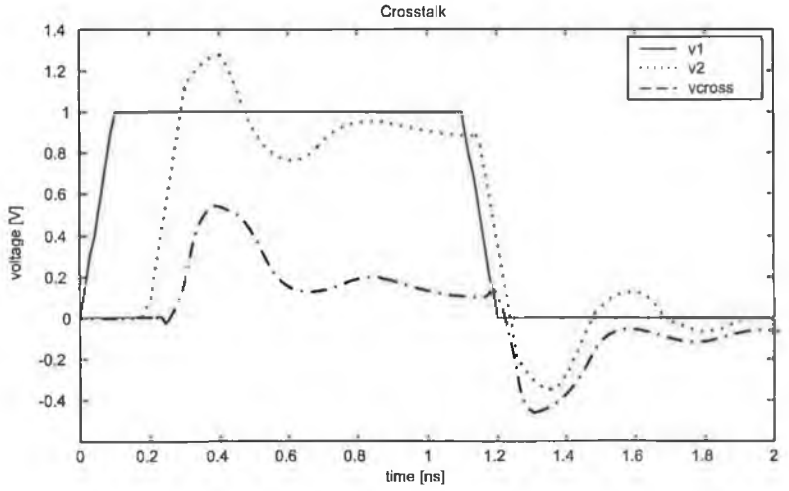
Modern compact and high-performing systems feature high levels of integration. This implies a consequent reduction in distances between signal lines. This leads to an increase in electromagnetic coupling through both mutual capacitance and inductance between neighbouring lines and unwanted interaction may occur. This interaction is termed *crosstalk* and by its very nature it involves a system of two or more conductors as shown in Fig 2.8.b).

The crosstalk arises when signal energy from the active line is coupled to the inactive line through both mutual capacitance and inductances, resulting in noise voltage and currents in the quiet line. For example, propagation of a signal in active line

1 in Fig 2.8.b) causes the appearance of a noise voltage $v_{crosstalk}$ on the adjacent line 2 which should be in its steady state since there are no active devices connected to the line. Obviously, such interaction may lead to all sorts of system glitches and has to be very carefully examined when designing high-speed circuits.



a) Input voltage *b) Network with multiconductor transmission line*



b) Time response

Fig. 2.8. *Illustration of crosstalk*

The crosstalk effect emerged as a serious problem for designs in the current 130nm technology and there are a lot of efforts to properly simulate the crosstalk effect. A simple linear RC circuit model, as proposed in [S92], is not satisfactory as shown by Deutsch et al. in [DKR+97]. Instead, the inclusion of the inductive effects is necessary [SHP+01], [DCK+01]. An extensive study of crosstalk simulation issues can be found in [DSS+99] as well as the guidelines as to when to use frequency dependant RLC models.

2.2.6. Current distribution effects

At higher frequencies, interconnects start to display behaviour that is dependent on the frequency of the signal propagating through it. This is mainly due to the fact that electromagnetic field penetration inside the conductor exhibits strong frequency dependence leading to *current distribution effects* [RC01], [AN01], [DS94], [D98], [YFW82]. At relatively low frequencies, the current in a conductor is distributed uniformly all over its cross section and the signal is propagating throughout all available area. The relationship between the line resistance and the conductor cross section area is:

$$R = \frac{\rho l}{A} \tag{2.6}$$

where ρ is volume resistance, l is line length and A is area of cross section. Thus, it is clear that the line resistance at low frequencies will be relatively low since the cross sectional area is the largest possible. With an increase of operating frequency, currents start to concentrate near the surface or edges of the conductor, i.e. the current distribution becomes uneven giving rise to effects such are skin effect, edge effect and proximity effect.

The *skin effect* causes the current to concentrate in a thin layer near the conductor surface and thus reduces the effective cross section area available for signal propagation. The measure of this uneven current distribution is known as *skin depth* δ and it is defined as the penetration distance at which current density is attenuated by 1 neper (1 neper = 1/e= -9.7dB) [D98]. The skin depth may be calculated as:

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} = \sqrt{\frac{\rho}{\pi\mu f}} \tag{2.7}$$

where ρ is relative electrical resistance, μ is magnetic permeability and $\omega = 2\pi f$ is the radian frequency of the propagating signal. As can be seen, with increasing frequency the skin depth is reduced thus leading to a reduction in the cross section area available for current flow. This in turn leads to an increase in the resistance to signal propagation and a decrease in inductance due to the decrease in the magnetic field inside the conductor. As a rule of thumb [D98], skin-effect occurs generally around the frequency where $\delta \leq 0.3t$, where t is the thickness of the conductor cross section.

The *edge effect* causes the current to concentrate near the sharp edges of the conductor. This, in turn, raises the total resistance of the line.

The *proximity effect* causes the current to concentrate in the sections of the ground plane that are close to the signal conductor. This leads to a decrease in line inductance but an increase in the line resistance.

Clearly, all current distribution related effects are heavily dependant on the frequency of the propagating signal and have to be described via frequency-dependant parameters. The variation of the interconnect electromagnetic properties with frequency is known as *dispersion* [RC01]. The frequency-dependence of the voltage/current ratio caused by dispersion, leads to considerable signal distortion at the high frequencies that has to be taken into account.

As seen in this section, there are various high-frequency interconnect effects and unfortunately, each has significant influence on the integrity of a propagating signal. Very often, advances in minimising one effect make another one more pronounced. For example, increasing the density of the circuit leads to shorter interconnects which reduces the problem of delay and reflections. But a higher density of wires leads to greater crosstalk problems and a more pronounced proximity effect. As the operational frequencies of the circuit move into the gigahertz range, these effects become more pronounced and have to be taken into account during the design stage. Hence there is a need for an efficient interconnect model that is capable of accurately capturing these high-frequency effects.

2.3. Electrical models of high-frequency interconnects

The first step in forming an electrical model of an interconnect is a mapping of the interconnect physical parameters such as length, cross-sectional dimension, dielectric and metal properties into appropriate electrical parameters (e.g. per-unit-length R, L, C, G). This process is called *extraction*. During synthesis of per-unit-length parameters causality constraints have to be enforced and met in order to ensure that the synthesized model accurately represents physical reality [CCH+01], [P98]. Although an accurate extraction process is an important issue in modelling interconnects, the process itself is beyond the scope of the research presented in this thesis. Hence, from this point forward it will be assumed that reasonably accurate electrical parameters are readily available for the interconnect structure for which an electrical model is sought. Depending on the operating frequency, signal rise times and the physical nature of structure, the interconnect model may be lumped, distributed, based on a tabulated data set or a full-wave model.

2.3.1. Lumped models

In principle, any line may be divided into several segments of smaller lengths. If the length of each of these segments is much smaller than the wavelength corresponding to the highest frequency of interest, than each segment may be replaced by a lumped model [P94], [P98]. The simplest model of an interconnect is the linear RC tree model, where each segment is modelled by a capacitor from the node to ground and a resistor in the direct branch as shown in Fig 2.9.

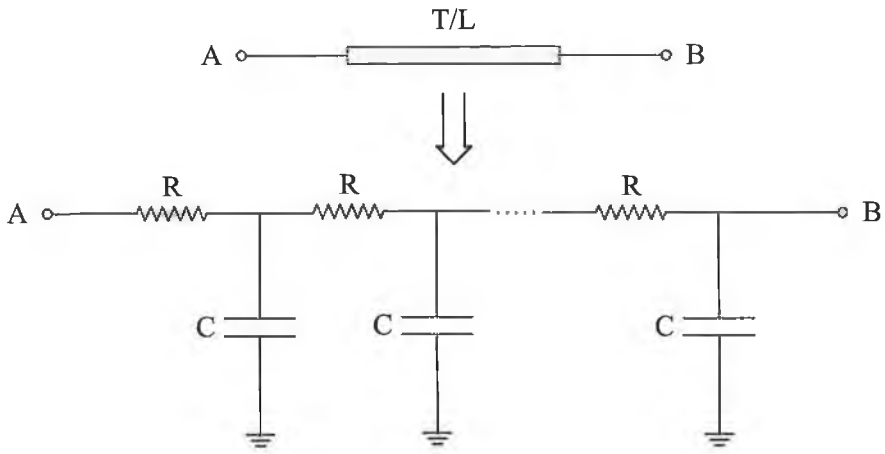


Fig. 2.9. *RC tree*

The linear RC model does not allow floating capacitors or resistors to the ground. Non-linear resistors and capacitors (e.g. for a MOS transistor) are approximated by a linear resistor and capacitor respectively. Such a tree has a monotonic response and thus is capable of predicting effects like propagation delay and attenuation without requiring significant CPU involvement.

But as the operating frequency increased, the ringing effect became more pronounced and its oscillating nature could not be successfully described by the monotonic response of an RC tree. It became necessary to include inductors into the model to allow a description of non-monotonic behaviour. In addition, grounded resistors were also introduced into the interconnect model yielding the general RLCG based lumped model shown in Fig 2.10.

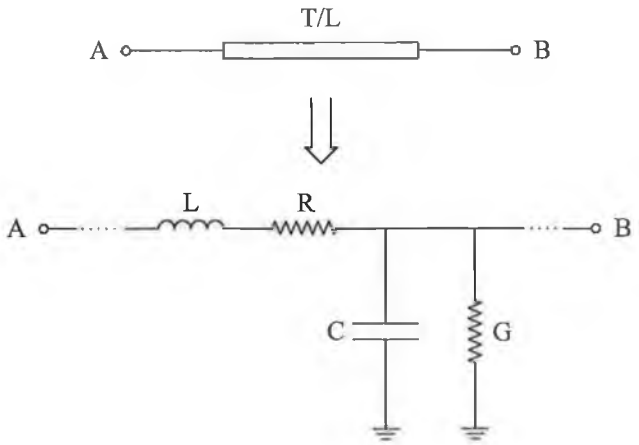


Fig. 2.10. *RLCG model*

There are several problems associated with a lumped model representation of a distributed structure. The biggest disadvantage of the lumped model is the very large number of sections necessary to correctly describe interconnect behaviour. For example, in the case of a lossless line described with conductance C and inductance L, the number of segments (N) required for reasonably accurate approximation of an interconnect is given by [NA02]:

$$N \geq \frac{10\tau l}{t_r} \tag{2.8}$$

where l is line length, $\tau = \sqrt{LC}$ is line delay and t_r is the rise time of a signal. As the circuit layout gets more and more complex with longer lines and delay times, and the rise time of a signal gets shorter, the number of sections necessary to take into account interconnect behaviour grows rapidly thus putting huge computational demands on circuit simulators.

An additional problem associated with a lumped model representation of a distributed structure described via the analytical wave equation model is related to the bandwidth. The lumped model representation is a discrete, finite-order approximation that is band limited. Such a representation will have an instantaneous response. On the other hand, distributed networks are of infinite-order and have inherent delay. Obviously such a delay cannot be properly modeled via a lumped model that is of finite-order [CCH+01]. One remedy for such a problem was suggested by Heeb and Ruehli [HR91] by including retardation, i.e. finite time delay, when the physical size of the elements is larger than 1/10 of a wavelength.

Furthermore, it is very difficult to deal with frequency-dependant parameters via a lumped model. In addition, the Gibbs phenomenon associated with a lumped model description introduces spurious ripples in the transient simulation results. Various lumped model representations are examined in detail by Deutsch et al. [DKR+97], [DCK+01]. The key deficiencies of a lumped model are highlighted and the use of a distributed model when simulating modern high speed interconnects is recommended.

2.3.2. Distributed transmission-line model

In a distributed model, an interconnect is modelled as a transmission line and TEM or quasi-TEM mode of signal propagation is assumed. The TEM mode is an ideal assumption where it is assumed that both the electrical (E) and magnetic (H) fields are perpendicular to the direction of wave propagation. It is valid under the condition that the line cross-section is much smaller then the wavelength. But in practice, there are always electrical and magnetic fields in the direction of propagation due to the interconnect inhomogeneities. So if, in the frequency range of interest, the line cross section or the extent of these non-uniformities remain a small fraction of the wavelength in the frequency range of interest, the solution of Maxwell's Equations describing interconnects is given by so-called quasi-TEM modes. Except in cases where abrupt discontinuities (vias, bends, etc.) are present, most interconnect will exhibit quasi-TEM behaviour [NA02], [D98].

In general, models based on the quasi-TEM approximation are characterised by distributed R, L, C, G per unit length (p.u.l.) parameters and the behaviour of the interconnect is described by a set of linear PDEs termed the Telegrapher's Equations:

$$\begin{aligned} \frac{\partial v(x,t)}{\partial x} &= -Ri(x,t) - L\frac{\partial i(x,t)}{\partial t} \\ \frac{\partial i(x,t)}{\partial x} &= -Gv(x,t) - C\frac{\partial v(x,t)}{\partial t} \end{aligned} \tag{2.9}$$

The Telegrapher's Equations represent voltages and currents in terms of both time and position along the interconnect and hence, the distributed nature of such a model.

The solution to (2.9) can be interpreted in terms of waves travelling forward and backward on the line [P94]. On the other hand, the phenomenon of wave propagation on transmission lines can be seen as an extension of circuit theory [111]. The key distinction between the circuit theory and transmission-line theory is the electrical wavelength λ. Circuit analysis is valid if the electrical wavelength is much greater than the physical dimensions of the network. In such a case, a lumped model representation

is valid and yields acceptable results. However, if the transmission line has physical dimensions comparable in size to the electrical wavelength, the transmission lines must be treated as a distributed-parameter network. Therefore, Paul [P94] suggests the use of the *electrical length* of the transmission line as a criterion in order to determine whether a lumped or distributed model is appropriate. If, at the highest operating frequency of interest, the interconnect length (d) is physically one order of size shorter than the wavelength (λ) (i.e. $\lambda/d \geq 10$), the interconnect is considered to be *"electrically short"* and can be represented with a lumped model since transmission line effects are negligible. However, as frequency increases, the corresponding wavelength decreases and becomes comparable to the length d, as can be seen from (2.10):

$$\downarrow \lambda = \frac{v}{f_{\uparrow}} \approx d\,. \tag{2.10}$$

In such a case, the interconnect is referred to as *"electrically long"* and needs to be modelled using distributed or full-wave models in order to capture interconnect effects which in that case have significant influence on overall signal propagation.

In general, the p.u.l. parameters may be a function of the distance along the line and/or are frequency dependant. For uniform lines, the p.u.l. parameters are constant with respect to distance along the line [P94]. However, owing to complex interconnect geometries and varying cross-sectional areas, modern interconnects often have to be modelled as non-uniform lines. In such a case, the per unit length parameters are a function of distance along the length of the transmission line [F93], [D98]. For correct delay and rise-time estimation, models based on frequency independent p.u.l. parameters are sufficient [DSS+99]. However, it has been shown in [DCK+01] that for accurate crosstalk, noise and delay in clock networks, a *frequency-dependant* distributed $R(f)L(f)C$ circuit representation is necessary. The frequency dependence of the distributed parameters is mainly due to the existence of the current distribution effects.

2.3.3. Models based on tabulated data

Modern interconnect networks often have a complex structure with non-uniform lines and other geometric inhomogeneities such as discontinuities (e.g. vias, bends, etc.) routinely present. For instance, interconnects in chip packages are usually non-uniform due to high circuit density, complex shapes and geometrical constraints. Obtaining an accurate analytical model for interconnect networks like this may be very difficult and sometimes impossible. To deal with such interconnects, an interconnect model based on

a tabulated set of data is utilised. This data may be from actual measurements or from electromagnetic simulations. A high-speed interconnect is then described in terms of frequency-dependent scattering (*s*), admittance (*y*), impedance (*z*) or hybrid (*h*) parameters. These parameters relate the terminal voltages and currents of a network. This is very convenient for most digital designers for whom the current and voltage distribution along the lines is of no interest unless electromagnetic interference (EMI) problems are being studied. Before measured/simulated data is included into an interconnect model based on this tabulated data set, usually a data preconditioning is necessary in order to ensure preservation of causality that may be lost due to measurement/simulation errors.

2.3.4. Full-wave model

With the ever-increasing operating frequency, the line cross section becomes a significant fraction of the wavelength and the field components in the direction of propagation can no longer be neglected. Consequently, two–dimensional (2-D) quasi-TEM distributed transmission-line models utilising the Telegrapher's Equations become inadequate to describe the spatial electromagnetic effects of three-dimensional (3-D) interconnect structures. In addition, electronically long interconnects can behave as spurious antennas and pick up emissions from other electronic equipment in close proximity, as well as radiating energy themselves [NA02]. The widespread use of wireless technologies and the high operational frequencies are emphasizing these electromagnetic interference (EMI) and electromagnetic compatibility (EMC) issues that cannot be effectively tackled without taking into account all field components.

In general, when the cross-sectional dimension reaches 1/10 of the effective wavelength, a full-wave model is needed to accurately describe interconnect effects. Additional matters, such as the ever-reducing distance between adjacent conductors and the increase in dielectric constant values should be considered when deciding on whether to use a full-wave model or distributed model. A full-wave model takes into account all possible field components and satisfies all boundary conditions of the corresponding Maxwell's Equations. An interconnect is then described in terms of modal parameters such as propagation constant and characteristic impedance. The model is highly accurate but computationally very costly to the extent that its use becomes practically prohibitive.

Another problem with the full-wave model is that the information provided by the model is not in terms of currents and voltages as required by standard circuit simulators. Hence, problems associated with combining the interconnect simulation results from a full-wave model with the rest of circuit is another serious issue. One remedy is the partial element equivalent structure (PEEC) model introduced by Ruehli [R74]. PEEC models are RLC circuits where individual resistances and capacitances are extracted from the geometry using a quasi-static (nonretarded) solution of Maxwell's Equations. Since the PEEC model is represented via RLC components it may be linked to a circuit simulator. For a more accurate full-wave solution, retarded PEEC (rPEEC) models [HR91] are used. PEEC models have demonstrated a high level of success when modelling interconnects that requires a 3-D simulation procedure. However (r)PEEC models result in large networks [R74], [RC01], [CRZ00] and simulation is very CPU intensive.

2.4. Interconnect simulation issues

Modern circuits are extremely complex and comprise of hundreds of thousands of interconnects and non-linear lumped elements. Simulation of such large systems is associated with two major problems: the mixed time/frequency nature of the simulation and the computational expense.

2.4.1. Mixed time/frequency domain

Including distributed interconnect models in a transient simulation in a general-purpose circuit simulator is very difficult. Circuit simulators such as SPICE [N75] are time-domain based since circuits containing devices with non-linear or time-dependent characteristics must be characterised in the time domain [BS97]. If the lumped RLCG model is sufficient to describe interconnect behaviour, a SPICE like simulator may be used for simulation purposes. This usually involves high CPU cost as SPICE does not handle large linear RLCG networks efficiently. Furthermore, as shown earlier, simple lumped models are inadequate to accurately describe the behaviour of modern high-speed interconnects and consequently, frequency dependent distributed models must be used raising the problem of mixed time/frequency domain.

The distributed models of interconnect are formulated in terms of time-domain partial differential equations (Telegrapher's Equations) but obtaining solution to them is very difficult if not impossible. However, in the frequency domain, the corresponding

description is a set of linear equations whose solution is straightforward to obtain. Additionally, if an interconnect has frequency-dependent parameters it is best described in the frequency domain [WW92] since dispersion, conduction and dielectric losses are relatively simple functions of frequency and are generally time invariant.

Therefore, in order to incorporate the transmission-line behaviour of interconnects into a general-purpose circuit simulator, it is necessary to convert frequency-domain results for interconnects into a time-domain description (Fig 2.11). Several approaches have been proposed in the literature, e.g. [XLW+00] and [B00].

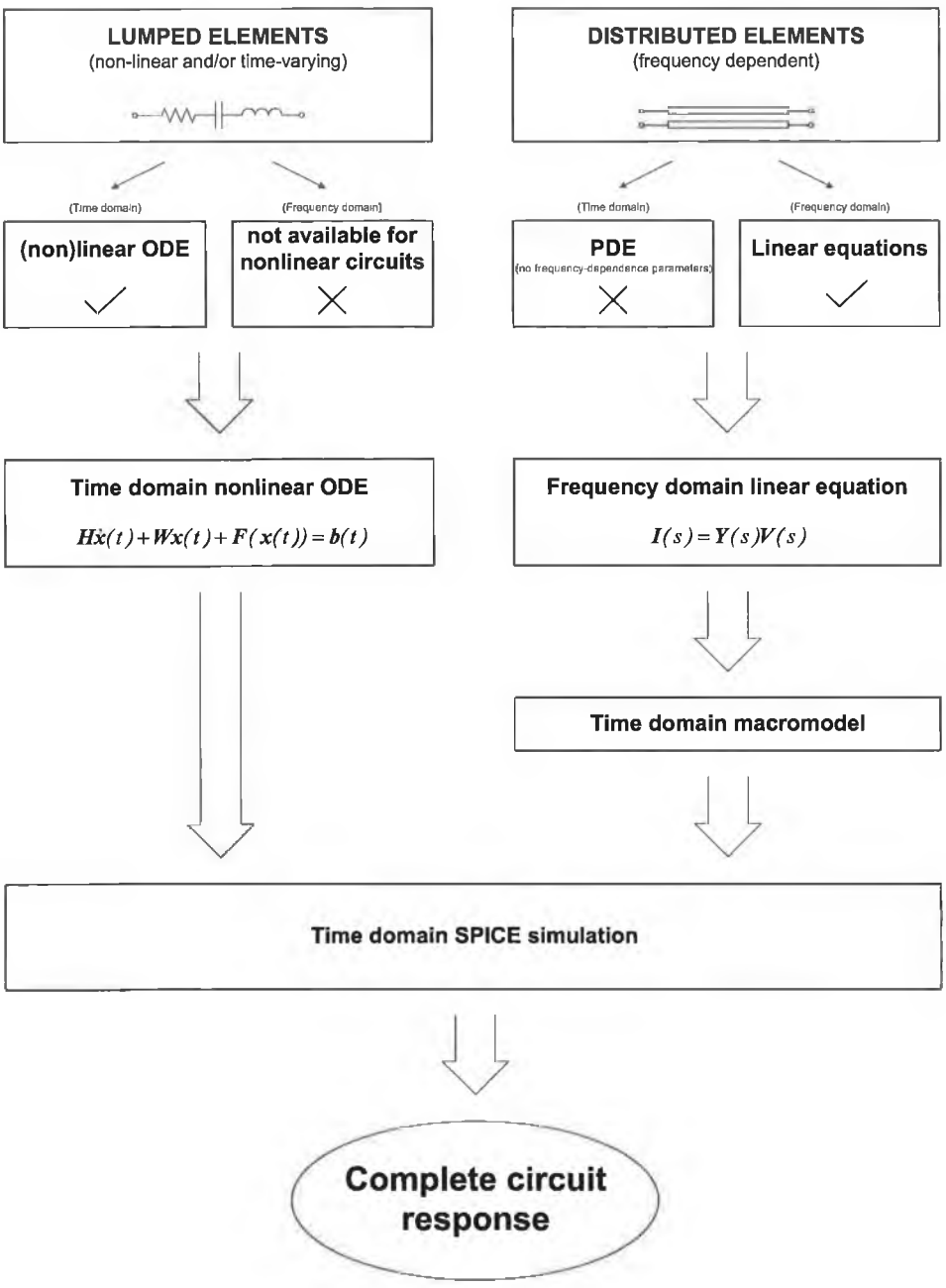


Fig. 2.11. *Mixed time/frequency domain problem*

2.4.2. Computational expense

The first step in the simulation process is to write a set of circuit equations that describe the circuit behaviour. These equations may be written either in the time-domain or in the frequency domain but, due to the mixed time/frequency issue, in the majority of cases the simulation has to be performed in the time domain. For the purpose of obtaining a numerical solution, integration techniques are used to convert a set of time-domain differential equations into a set of difference equations. Then the Newton iteration process is applied in order to obtain simulation results at each time point. However, the matrices that ensue from the set of difference equations describing the interconnect network are usually very large and thus LU decompositions performed as part of the Newton algorithm place a heavy demand on CPU processing time. Additionally, memory requirements may be overwhelming for large networks. To address this problem, model order reduction techniques are introduced. They enable a speed up of calculations but introduce new problems regarding ill-conditioning of large matrices and preservation of the stability and passivity of the reduced model.

2.5. Summary

As VLSI feature sizes reach deep sub-micron dimensions and clock frequencies approach the gigahertz range, interconnect effects such as propagation delay, attenuation, crosstalk, signal reflection, ringing and current distribution effects become an increasingly significant factor in determining overall system performance. Hence, the ability to describe high-frequency interconnect effects in an effective and accurate manner is a must for any state-of-art interconnect model.

An interconnect model can be a lumped model (RC or RLCG), a distributed model (with or without frequency-dependent parameters), a model based on a tabulated data set or a full-wave model. The interconnect length, cross-sectional dimensions, signal rise time and the clock speed are factors which should be examined when deciding on the type of model to be used for modelling high-speed interconnects. In addition, it might be necessary to take into consideration other factors such as logic levels, dielectric materials and conductor resistance.

With the trend of ever-rising operational frequencies and ever-shrinking feature sizes, lumped models became insufficient to adequately describe the behaviour of modern high-speed interconnects. The full-wave model, although very accurate, is too

computationally involved and cannot produce simulation results in a reasonable amount of time. Therefore, this thesis will focus on distributed interconnect models described in terms of the Telegrapher's Equations and models based on a tabulated data set. The aim is to obtain interconnect models that are capable of describing non-uniform and frequency-dependant interconnects with reasonable accuracy and in a computationally efficient manner.

CHAPTER 3

Interconnect Simulation Techniques

Except for very simple interconnect networks and structures (e.g. short lossless lines), accurate simulation of interconnects is not a simple task. SPICE-like simulators cannot handle the large numbers of state variables associated with the description of an interconnect in terms lumped resistors, inductors and capacitors [CC98]. In particular, the extensive mutual inductive and capacitive coupling present in the equivalent model, makes SPICE-based simulation prohibitively slow if at all possible [CCP+98]. Therefore, during the last twenty years, substantial research into developing accurate and efficient techniques for modelling and simulation of interconnects has been carried out. The resulting *interconnect simulation techniques* can be broadly classified into two main categories [AN01]: approaches based on macromodelling of each individual transmission line set and approaches based on model order reduction (MOR) of the entire linear network containing both lumped and distributed subnetworks.

The goal of this Chapter is to review some of the existing *interconnect simulation techniques* and highlight their merits and demerits. The basic properties of a distributed network are first introduced followed by a short description of the most widely used macromodelling and model order reduction strategies.

3.1. An overview of distributed network theory

As explained in Chapter 2, assuming TEM or quasi-TEM mode of propagation along the line, interconnect behaviour may be characterised by the Telegrapher's Equations. In this section, some basic properties relevant to networks described by the Telegrapher's Equations are introduced [P94], [P98].

3.1.1. Time-domain Telegrapher's Equations

In order to analyse the distributed line, the standard approach is to discretise the line under consideration into infinitely small sections of length Δx. According to quasi-static field theory, the voltage drop along this length Δx is the overall result of magnetic

(impedance) couplings. The change in current over a length Δx is the overall sum of capacitive currents associated with the electrostatic field related to the voltage distribution on the line. Therefore, the behaviour of the section of the distributed line of length Δx may be approximated by a lumped-element equivalent circuit shown in Fig. 3.1 comprising of inductance, capacitive and resistive elements. A finite length of distributed line can be viewed as a cascade of such sections and its behaviour is characterised by the Telegrapher's Equations.

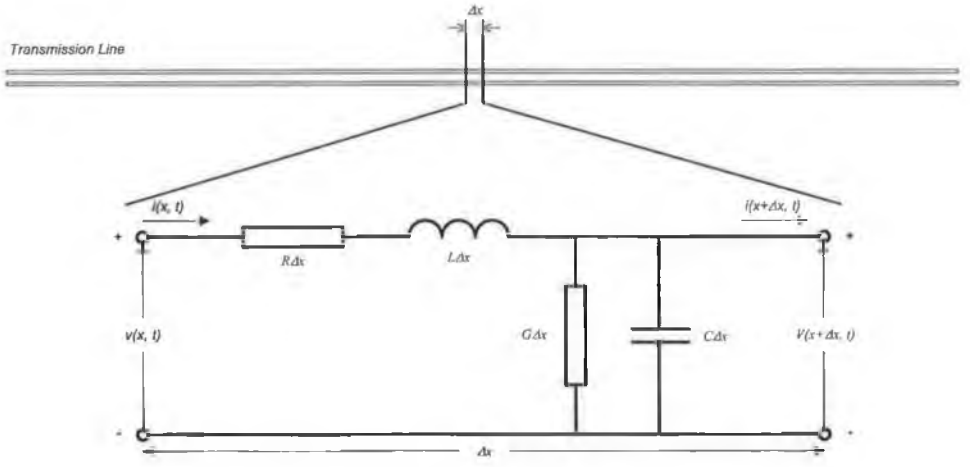


Fig. 3.1. *Lumped-element equivalent circuit*

In order to derive the Telegrapher's Equations, Kirchhoff's voltage law is first applied to the circuit in Fig. 3.1 yielding:

$$v(x+\Delta x,t)=v(x,t)-R\Delta x i(x,t)-L\Delta x\frac{\partial i(x,t)}{\partial t}. \tag{3.1}$$

Kirchhoff's current law applied to the same circuit gives:

$$i(x+\Delta x,t)=i(x,t)-G\Delta x v(x+\Delta x,t)-C\Delta x\frac{\partial v(x+\Delta x,t)}{\partial t}. \tag{3.2}$$

Dividing (3.1) and (3.2) by Δx and taking the limit as $\Delta x \to 0$ gives the following differential equations:

$$\begin{aligned}\frac{\partial v(x,t)}{\partial x}&=-Ri(x,t)-L\frac{\partial i(x,t)}{\partial t}\\ \frac{\partial i(x,t)}{\partial x}&=-Gv(x,t)-C\frac{\partial v(x,t)}{\partial t}\end{aligned} \tag{3.3}$$

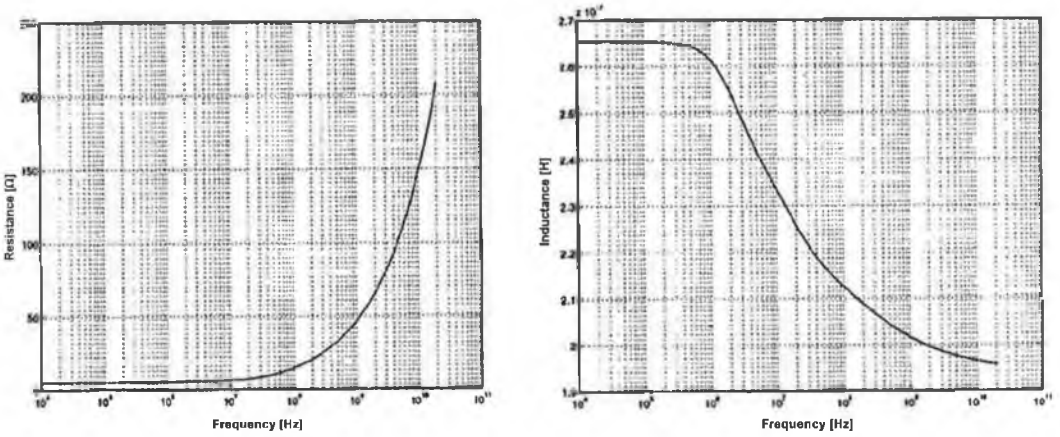
Equations (3.3) are the *time-domain* form of the *Telegrapher's Equations*. They are a set of linear PDEs describing voltages and currents in terms of both time and position along the transmission line.

The distributed nature of a transmission line is typically described by *per-unit-length (p.u.l.)* parameters (R, G, L and C) defined on a lumped-element equivalent

circuit, shown in Fig 3.1, for a short piece of the line of length Δx. The *series resistance p.u.l. R* [Ω/m] represents the resistance due to the finite conductivity of the conductors and the *shunt conductance p.u.l. G* [S/m] is due to dielectric loss in the material between the conductors. The *series inductance p.u.l. L* [H/m] represents the total self-inductance of the two conductors and the *shunt capacitance p.u.l. C* [F/m] is due to the close proximity of the two conductors. P.u.l. parameters are usually extracted over a certain frequency range of interest from either measured data or results from a full-wave simulation.

3.1.2. Frequency dependant p.u.l. parameters

At relatively low frequencies, current distribution effects are negligible and the value of the p.u.l. parameters remains constant with respect to frequency. However, at high- and mid-frequency ranges, current distribution effects can cause significant changes in the values of the resistance and inductance p.u.l. parameters as illustrated in Fig. 3.2 [AN01].



a) Resistance *b) Inductance*

Fig. 3.2. *Illustration of frequency dependence of resistance and inductance*

As can be seen, p.u.l. resistance is a relatively small constant value at low- and mid-frequency ranges. But at the high frequencies, the increase of resistance with the increase in frequency is exponential in nature and has to be accounted for during the design process. On the other hand, p.u.l. inductance is a relatively high constant value at low frequencies and drops considerably throughout the mid-range to become again a constant value at high frequencies. P.u.l. capacitance remains more or less constant, since it is mostly a function of geometry and is not influenced by frequency [DCK+01], while p.u.l. conductance is mostly influenced by the frequency dependence of the so

called loss tangent [RC01] defined by $tan\delta = \sigma / \omega\varepsilon$ where σ is conductivity of the line, $\omega = 2\pi f$ is the radial frequency of propagating signal and ε is the dielectric constant of the medium. If it is deemed necessary to take into account these changes in the line's parameters with operating frequency, a high-speed interconnect has to be modelled with frequency dependant per-unit-length parameters.

3.1.3. Frequency-domain Telegrapher's Equations

Taking the Laplace transform of (3.3) with respect to time, one can write the following *Laplace domain* form of the Telegrapher's Equations:

$$\begin{aligned} \frac{dV(x,s)}{dx} &= -(R+sL)I(x,s) = -ZI(x,s) \\ \frac{dI(x,s)}{dx} &= -(G+sC)V(x,s) = -YV(x,s) \end{aligned}, \tag{3.4}$$

where $Z = R + sL$ is the *p.u.l. impedance* of a transmission line and $Y = G + sC$ is the *p.u.l. admittance* of a transmission line. In general, both Z and Y are dependent on position along the line, i.e. $Z = Z(x)$ and $Y = Y(x)$. Setting $s = j\omega$, the *frequency domain* Telegrapher's Equations are obtained as:

$$\begin{aligned} \frac{dV(x,\omega)}{dx} &= -(R+j\omega L)I(x,\omega) = -ZI(x,\omega) \\ \frac{dI(x,\omega)}{dx} &= -(G+j\omega C)V(x,\omega) = -YV(x,\omega) \end{aligned}. \tag{3.5}$$

As can be seen, a set of time-domain PDEs (3.3) is now converted to a set of ODEs involving variations of voltages and currents with respect to distance at a given frequency ω. Bearing this in mind (3.5) may be compactly noted as:

$$\frac{dV(x)}{dx} = -Z(x)I(x) \tag{3.6}$$

$$\frac{dI(x)}{dx} = -Y(x)V(x). \tag{3.7}$$

3.1.4. Uniform lines

Assuming that the transmission line is *uniform* (at least over defined lengths), Z and Y are independent of the distance parameter x so that (3.6) can be differentiated as:

$$\frac{d^2V(x)}{dx^2} = -Z\frac{dI(x)}{dx}. \tag{3.8}$$

Direct inward substitution, using (3.7) to eliminate $I(x)$, then gives the second order ODE from which the voltage $V(x)$ may be calculated

$$\frac{d^2V(x)}{dx^2} = ZYV(x). \tag{3.9}$$

An analogous equation for obtaining the current $I(x)$ may be obtained as

$$\frac{d^2I(x)}{dx^2} = YZI(x). \tag{3.10}$$

Defining the *complex propagation constant* γ as

$$\gamma = \sqrt{ZY} \tag{3.11}$$

equations (3.9) and (3.10) may be compactly written as:

$$\begin{aligned} \frac{d^2}{dx^2}V(x) &= \gamma^2 V(x) \\ \frac{d^2}{dx^2}I(x) &= \gamma^2 I(x) \end{aligned} \tag{3.12}$$

The complex propagation constant γ is a function of frequency and may be noted as:

$$\gamma = \sqrt{ZY} = \sqrt{(R+j\omega L)(G+j\omega C)} = \alpha(\omega) + j\beta(\omega) \tag{3.13}$$

where α is the *attenuation constant* given in nepers/m and β is the *phase constant* given radians/m.

It is well known that the solution to (3.12) may be written as a combination of waves travelling forward and backward on the line as:

$$\begin{aligned} V(x) &= V_0^+ e^{-\gamma x} + V_0^- e^{+\gamma x} \\ I(x) &= I_0^+ e^{-\gamma x} + I_0^- e^{+\gamma x} \end{aligned} \tag{3.14}$$

where the $e^{-\gamma x}$ term represents wave propagation in the $+x$ direction and the $e^{+\gamma x}$ term represents wave propagation in $-x$ direction. The phase shift experienced by the travelling waves is given by $e^{\pm j\beta(s)x}$ and attenuation is characterised by $e^{\pm\alpha(s)x}$. Equations (3.14) are referred to as the *travelling wave solution* to the Telegrapher's Equation.

A *characteristic impedance*, Z_0, of a transmission line is defined as:

$$Z_0 = \frac{R+j\omega L}{\gamma} = \sqrt{\frac{R+j\omega L}{G+j\omega C}} = \sqrt{\frac{Z}{Y}} \tag{3.15}$$

and the relationship between the amplitudes of forward/backward travelling voltage and current waves is given as:

$$\frac{V_0^+}{I_0^+} = Z_0 = -\frac{V_0^-}{I_0^-}. \tag{3.16}$$

For interconnect structures, the value of Z_0 is in the range of 30 – 60 Ω, and most on-chip interconnects have Z_0 in the range of 45-55 Ω [DCK+01]. Consequently, most designers use Z_0 = 50Ω as a good first approximation without performing a time-costly full analysis of the interconnect structure in question.

The *wavelength* on the line, λ, is defined as:

$$\lambda = \frac{2\pi}{\beta} \tag{3.17}$$

and the *phase velocity*, v_p, is

$$v_p(\omega) = \frac{\omega}{\beta} = \lambda f = \frac{\omega}{Im\{\sqrt{(R+j\omega L)(G+j\omega C)}\}} \tag{3.18}$$

As can be seen from (3.15), (3.17) and (3.18), in the general case, the characteristic impedance Z_0, the wavelength λ and the phase velocity v_p are functions of frequency ω.

In some practical cases, in the low– and mid–frequency range, the losses of the line represented by R and G are very small and may be neglected, i.e. $R = G = 0$. Such a line is then called *lossless*. When losses cannot be neglected ($R \neq 0,\ G \neq 0$), the line is termed *lossy* line. For lossless lines, the attenuation constant α is zero and the transmission line represents a pure-delay element. The characteristic impedance Z_0 becomes a purely real number and is not dependant on frequency. In addition, the phase velocity for a lossless line is also independent of frequency ω.

3.1.5. Multiconductor transmission line (MTL) systems

In practical applications, a single transmission line (STL) system as given in Fig 3.3 is rarely found. Instead, a multiconductor transmission line (MTL) system with N coupled conductors shown in Fig 3.4 is the norm [AN01].

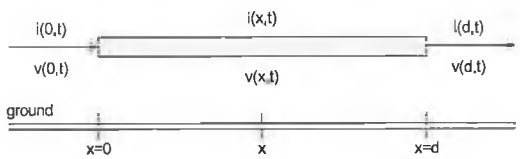


Fig. 3.3. *Single transmission line (STL) system*

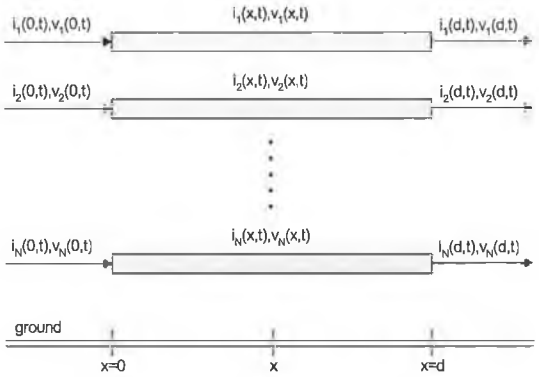


Fig. 3.4. *Multiconductor transmission line (MTL) system*

The set of equations describing a MTL system in the time–domain analogue to (3.3) may be written as a set of *2N* coupled first-order PDEs [P94]:

$$\begin{aligned} \frac{\partial \boldsymbol{v}(x,t)}{\partial x} &= -\boldsymbol{R}\boldsymbol{i}(x,t) - \boldsymbol{L}\frac{\partial \boldsymbol{i}(x,t)}{\partial t} \\ \frac{\partial \boldsymbol{i}(x,t)}{\partial x} &= -\boldsymbol{G}\boldsymbol{v}(x,t) - \boldsymbol{C}\frac{\partial \boldsymbol{v}(x,t)}{\partial t} \end{aligned}, \tag{3.19}$$

or in matrix form as:

$$\frac{\partial}{\partial x}\begin{bmatrix} \boldsymbol{v}(x,t) \\ \boldsymbol{i}(x,t) \end{bmatrix} = -\begin{bmatrix} 0 & \boldsymbol{R} \\ \boldsymbol{G} & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{v}(x,t) \\ \boldsymbol{i}(x,t) \end{bmatrix} - \begin{bmatrix} 0 & \boldsymbol{L} \\ \boldsymbol{C} & 0 \end{bmatrix}\frac{\partial}{\partial x}\left(\begin{bmatrix} \boldsymbol{v}(x,t) \\ \boldsymbol{i}(x,t) \end{bmatrix}\right). \tag{3.20}$$

P.u.l. parameters become matrices (***R***, ***L***, ***G*** and ***C***) and the voltage/current variables become vectors ***v*** and ***i*** respectively. Symmetric and positive definite [P94], [NA02] matrices ***R***, ***L***, ***G*** and ***C*** are obtained by a 2-D solution of Maxwell's Equations along the transmission line using techniques based on a quasi-static or full-wave approach depending on the required accuracy and the geometry and structure of the line in question.

In the frequency domain, equations (3.4) become:

$$\begin{aligned} \frac{\partial \boldsymbol{V}(x,s)}{\partial x} &= -(\boldsymbol{R}+s\boldsymbol{L})\boldsymbol{I}(x,s) = -\boldsymbol{Z}\boldsymbol{I}(x,s) \\ \frac{\partial \boldsymbol{I}(x,s)}{\partial x} &= -(\boldsymbol{G}+s\boldsymbol{C})\boldsymbol{V}(x,s) = -\boldsymbol{Y}\boldsymbol{V}(x,s) \end{aligned}, \tag{3.21}$$

where $\boldsymbol{V}(x)$ and $\boldsymbol{I}(x)$ are vectors of line voltages and currents whose dimension is equal to the number of active lines. The earth's return path is taken as the reference for convenience. ***Z*** and ***Y*** are now impedance and admittance matrices given by:

$$\boldsymbol{Z} = \boldsymbol{R} + s\boldsymbol{L}, \quad \boldsymbol{Y} = \boldsymbol{G} + s\boldsymbol{C}\,. \tag{3.22}$$

The solution to (3.21) may be interpreted as corresponding to wave propagation [P94]. Natural modes of wave-propagation for a general multiconductor system may be obtained by diagonalising ***ZY*** [W63].

As can be seen, the equations that describe the behaviour of a MTL (3.19) are analogues to the equations for a STL (3.3). Hence, in many practical cases, the techniques developed for analysing a STL many readily be extended to describe MTL system behaviour.

3.2. Strategies based on transmission line macromodelling

The common property of most *interconnect macromodelling strategies* is that they introduce some kind of discretization of the set of partial differential equations that describe the interconnect network (Telegrapher's Equations). The result of this discretization is a set of ordinary differential equations called the *macromodel*. Then the macromodel equations may be linked into a circuit simulator and solved with a built-in ODE solver to obtain the overall response of a circuit. In the rest of this section, a brief review of the most representative macromodelling techniques is given.

3.2.1. Lumped segmentation technique

The lumped segmentation technique is the simplest approach that follows directly from the lumped-element equivalent circuit shown in Fig 3.1. In order to obtain a numerical solution to the Telegraphers Equations (3.3), the line of length l is divided into N smaller segments of the finite length Δx [P94]. If Δx is chosen such that it is electrically small at the frequencies of interest ($\Delta x \ll \lambda$), then each segment may be represented by a lumped-element equivalent circuit comprising of series elements $L\Delta x$ and $R\Delta x$, and shunt elements $G\Delta x$ and $C\Delta x$ as shown in Fig 3.1. Introducing this lumped interconnect representation into a circuit simulator is then a straightforward task.

However, the choice of appropriate Δx represents a major difficulty in a practical implementation of this technique as it depends both on the rise/fall time of the propagating signal (the pulse bandwidth) and the electrical length of the interconnect [CPP+99]. As a simple example, in order to accurately represent a lossless line of length l by LC segments, N needs to be at least [AN01]

$$N = \frac{10l\sqrt{LC}}{t_r} \tag{3.23}$$

where t_r is the rise time of signal. For a lossless line of $l=10\ cm$ and rise time of $0.2ns$ with p.u.l. parameters of $L=5\ nH/cm$ and $C=1\ pF/cm$, the number of segments required is $N \approx 35$. If losses are to be taken into account this number is even higher, i.e. for accurate simulation of GHz signals the number of segments per minimum wavelength is 15-20 [CC98]. Clearly, the size of such a model involves extremely long simulation times and huge memory requirements. In addition, direct lumped segmentation is insufficient to accurately describe frequency-dependent lines. Furthermore, the associated Gibbs phenomenon leads to ringing in the waveform that cannot be

completely eliminated from the waveform regardless of the number of segments utilised. Therefore, the direct lumped segmentation technique is not appropriate for modern high-speed interconnect modelling.

3.2.2. Direct time-stepping scheme

Lee, et al. [LKS93] suggest a direct time stepping scheme based on the finite element method. At each time step, a one dimensional boundary value problem is solved and values for the currents and voltages associated with the element are obtained as:

$$\begin{aligned} I_j^{n+\frac{1}{2}} &= \frac{2L-R\Delta t}{2L+R\Delta t} I_j^{n-\frac{1}{2}} - \frac{2\Delta t}{(2L+R\Delta t)\Delta x}(V_{i+1}^n - V_i^n) \\ V_i^{n+1} &= \frac{2C-G\Delta t}{2C+G\Delta t} V_i^n - \frac{2\Delta t}{(2C+G\Delta t)\Delta x}(I_{j+1}^{n+\frac{1}{2}} - I_j^{n+\frac{1}{2}}) \end{aligned} \qquad (3.24)$$

As can be seen, the value for current is computed one half time step before the value for the voltage in a so called leap-frog scheme. However, for simulation of high-frequency interconnects the time-step, Δt, would have to be extremely small in order to capture the fast transients that occur on the line. Hence, the CPU expense associated with the direct time-stepping scheme is unacceptably high. Therefore, the direct time-stepping algorithms are not recommended for use for simulation of high-frequency interconnects.

3.2.3. Convolution techniques

Djordjevic, et al. [DSB86] proposed a convolution approach for simulating interconnects exploiting the fact that an interconnect represents a linear system. It is well known that an output of a linear system, $y(t)$, may be expressed as a convolution of its input, $x(t)$, with the impulse-response of a system, $h(t)$, as [OWH+96]:

$$y(t) = \int_0^t x(\tau)h(t-\tau)d\tau \qquad (3.25)$$

Assuming that $x(t)$ is a piecewise-linear function, the numerical solution to this integral at a discrete time-point t_n may be obtained as [RP91]:

$$\begin{aligned} y(t_n) &= \int_0^{t_n} x(\tau)h(t_n-\tau)d\tau \\ &\approx x_n \frac{F(t_n-t_{n-1})}{t_n-t_{n-1}} + \sum_{i=1}^{n-1} x_i \left[\frac{F(t_n-t_{i-1})-F(t_n-t_i)}{t_i-t_{i-1}} - \frac{F(t_n-t_i)-F(t_n-t_{i+1})}{t_{i+1}-t_i} \right] \end{aligned} \qquad (3.26)$$

where $x_i = x(t_i)$ and $F(t) = \int_0^t \int_0^\tau h(\tau')d\tau' d\tau$.

Several other techniques use the convolution-based technique in combination with a Fast Fourier Transform (FFT) [DS87], recursive formulas [LK92] and state-space approaches [RNP94]. However, all of these techniques suffer from a common drawback. As can be seen from (3.26), numerical convolution requires integration over past history and thus is extremely computationally intense. Although the recursive formulation reduces the computational cost, it is still relatively high.

3.2.4. The method of characteristics (MC)

The method of characteristics (MC), introduced by Branin [B67], transforms a PDE representation of the lossless transmission line into an ODE along characteristic lines. An arbitrary lossless transmission line can be modelled by two impedances and two voltage controlled sources with time delay in the time domain enabling an easy linkage to transient simulators. In essence, time-delayed controlled sources extract the pure delay on the line, and "delayless" terms are then approximated with rational functions. Therefore, the MC is especially suitable for long low loss lines where the signal delay is pronounced. However, for n coupled lines, the MC requires $(2n^2+n)$ transfer functions [NA04] thus making the MC very computationally expensive. Furthermore, the MC macromodel cannot guarantee passivity.

Chang [C89] combined the MC with the waveform relaxation technique and Padé synthesis of the characteristic impedance and the complex propagation constant yielding the *generalized* MC that can deal with lossy coupled transmission lines. This method avoids time-domain convolution by solving the line equations in frequency domain. However, the computational efficiency is drastically reduced when compared to the MC for the lossless case since an FFT is used to transform the result back and forth between the time and frequency domain at each iteration. When high-speed interconnect is considered, a large number of data points is necessary to avoid the aliasing associated with the FFT. Xu, et al. [XLW+00] recently introduced a *modified* MC for analysis of uniform lossy lines where the characteristic admittance is modelled via a Taylor approximation and a Padé approximation is used to model the propagation constant. The application of the modified MC is limited to uniform lines and the passivity of the model is guaranteed.

3.2.5. Exponential matrix-rational approximation (EMRA)

The exponential Padé-based matrix-rational approximation (EMRA) uses Padé rational approximation of exponential matrices to convert PDEs into a time-domain set of ODEs [DNA99]. Consider the exponential form of the Telegrapher's Equations describing the multiconductor transmission line:

$$\begin{bmatrix} \boldsymbol{V}(l,s) \\ \boldsymbol{I}(l,s) \end{bmatrix} = e^{-\boldsymbol{Z}l} \begin{bmatrix} \boldsymbol{V}(0,s) \\ \boldsymbol{I}(0,s) \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} 0 & \boldsymbol{R}+s\boldsymbol{L} \\ \boldsymbol{G}+s\boldsymbol{C} & 0 \end{bmatrix} \tag{3.27}$$

where l is the length of the line. Matrix $e^{\boldsymbol{X}}$ may be approximated as

$$\boldsymbol{P}_{N,M}(\boldsymbol{X})e^{\boldsymbol{X}} \approx \boldsymbol{Q}_{N,M}(\boldsymbol{X}) \tag{3.28}$$

where $\boldsymbol{P}_{N,M}(\boldsymbol{X})$ and $\boldsymbol{Q}_{N,M}(\boldsymbol{X})$ are polynomial matrices expressed in terms of closed-form Padé rational functions. Setting $\boldsymbol{X} = -\boldsymbol{Z}l$ and after some mathematical manipulation a macromodel represented by a set of ODEs may be obtained [DNA99]. Since all the coefficients describing the macromodel are computed *a priori* and analytically, the method does not suffer from the usual ill-conditioning that is characteristic for direct use of Padé approximation. It may be proven that the EMRA algorithm preserves passivity [DNA99]. The computational advantage of the algorithm is obvious [AN01] and the EMRA provides fast models for shorter lines (e.g. on-chip wiring and board wires). However, the EMRA method is not well suited for the long, relatively lossless lines (e.g. several meters long coaxial cables) and the MC approach outperforms it due to its capability to extract the line delay that is the most significant factor for the performance of the long line [EHR+02].

3.2.6. Basis function approximation

Basis function approximation aims to express the variations in space for voltages and currents in terms of known basis functions, such as Chebyshev polynomials [CC97] or wavelets [BR00], [GC01]. For example, the voltage, $v(x,t)$, and current, $i(x,t)$, and their derivatives may be expanded in the form:

$$\begin{aligned} v(x,t) &= \sum_{n=0}^{N} a_n(t)F_n(x), & \frac{\partial}{\partial x}v(x,t) &= \sum_{n=0}^{N} \hat{a}_n(t)F_n(x) \\ i(x,t) &= \sum_{n=0}^{N} b_n(t)F_n(x), & \frac{\partial}{\partial x}i(x,t) &= \sum_{n=0}^{N} \hat{b}_n(t)F_n(x) \end{aligned} \tag{3.29}$$

where the coefficients $a_n(t)$, $b_n(t)$, $\hat{a}_n(t)$ and $\hat{b}_n(t)$ are now unknown variables. Functions $F_n(x)$ are functions chosen in such a manner as to form an orthogonal basis. Coefficients $a_n(t)$ and $b_n(t)$ are related to $\hat{a}_n(t)$ and $\hat{b}_n(t)$ as [AN01]:

$$a_n(t) = \frac{1}{2n}(\hat{a}_{n-1}(t) - \hat{a}_{n+1}(t))$$
$$b_n(t) = \frac{1}{2n}(\hat{b}_{n-1}(t) - \hat{b}_{n+1}(t)) \tag{3.30}$$

By substituting (3.29) into (3.3) and using the orthogonal properties of basis functions, the Telegrapher's Equations are converted to a set of ODEs in terms of the unknowns $a_n(t)$ and $b_n(t)$. A standard ODE solver may then be applied to obtain the line's response.

The advantage of this approach is that it is more computationally efficient than direct lumped RLC segmentation and that it can be readily applied to interconnects with non-uniform line parameters. The drawback is that when this algorithm is used with model order reduction model, passivity cannot be guaranteed [CC98], [AN01].

3.2.7. Compact-finite-differences approximation

The compact-finite-difference approximation method [CPP+99] also expresses the variations in space of the voltages $V(x,s)$ and currents $I(x,s)$ on a transmission line in terms of known expansion functions. However, it does so in the frequency domain. The spatial derivatives of $V(x,s)$ and $I(x,s)$ are approximated using the central difference operator

$$\alpha_i \left.\frac{\partial f(x)}{\partial x}\right|_{i+1} + \alpha_2 \left.\frac{\partial f(x)}{\partial x}\right|_{i} + \alpha_1 \left.\frac{\partial f(x)}{\partial x}\right|_{i-1} = \frac{f_{\frac{i+1}{2}} - f_{\frac{i-1}{2}}}{\Delta x} \tag{3.31}$$

where i denotes the node where the operator is centred and $f(x)$ represents either $V(x)$ or $I(x)$. The coefficients α_1 and α_2 are obtained such that the truncation error criteria are satisfied. The advantage of this algorithm is that achieves better accuracy with fewer variables than direct lumped segmentation and the passivity of the macromodel is guaranteed by construction [CPP+99], [AN01].

3.2.8. Integrated congruence transform (ICT)

The congruence transform approach, as introduced by Kerns, et al. [KWY95], guarantees the passivity of the RC based interconnect model only. However, it has been

extended to incorporate general RLC circuits [KY97], [OC97], [EL97]. In order to deal with distributed network modelling Yu, et al. [YWK99] established the integrated congruence transform (ICT). In the ICT, each distributed line is modelled by a finite order system with passivity preservation and explicit multipoint moment matching of its input admittance/impedance matrix. The Laplace domain equations (3.21) are first rewritten in the form:

$$\left(s\boldsymbol{M}(x)+\boldsymbol{N}(x)+\boldsymbol{T}\frac{d}{dx}\right)\boldsymbol{Z}(x,s)=0 \tag{3.32}$$

where

$$\begin{aligned}\boldsymbol{Z}(x,s)&=\begin{bmatrix}\boldsymbol{I}(x,s)\\ \boldsymbol{V}(x,s)\end{bmatrix}, \quad \boldsymbol{M}=\begin{bmatrix}\boldsymbol{L} & 0\\ 0 & \boldsymbol{C}\end{bmatrix},\\ \boldsymbol{N}&=\begin{bmatrix}\boldsymbol{R} & 0\\ 0 & \boldsymbol{G}\end{bmatrix}, \quad \boldsymbol{T}=\begin{bmatrix}0 & I\\ I & 0\end{bmatrix}.\end{aligned} \tag{3.33}$$

Then the following transform

$$\boldsymbol{Z}(x,s)=\boldsymbol{u}(x)\hat{z}(s) \tag{3.34}$$

is introduced, where transformation matrix $\boldsymbol{u}(x)$ is a function of spatial dimension only. Substituting (3.34) into (3.32), multiplying by $\boldsymbol{u}^T(x)$ and integrating with the respect to the normalised variable x, one obtains following equation

$$\left(s\hat{\boldsymbol{M}}+\hat{\boldsymbol{N}}+\hat{\boldsymbol{T}}\right)\hat{z}(s)=0\,, \tag{3.35}$$

where $\hat{\boldsymbol{M}}$, $\hat{\boldsymbol{N}}$ and $\hat{\boldsymbol{T}}$ are defined as:

$$\begin{aligned}\hat{\boldsymbol{M}}&=\int_0^l \boldsymbol{u}^T(x)\boldsymbol{M}(x)\boldsymbol{u}(x)dx,\\ \hat{\boldsymbol{N}}&=\int_0^l \boldsymbol{u}^T(x)\boldsymbol{N}(x)\boldsymbol{u}(x)dx,\\ \hat{\boldsymbol{T}}&=\int_0^l \boldsymbol{u}^T(x)\boldsymbol{T}\frac{d\boldsymbol{u}(x)}{dx}dx.\end{aligned} \tag{3.36}$$

After some mathematical manipulations [YWK99], equations (3.36) can be translated to a set of ODEs that form the macromodel. The macromodel formed via ICT preserves passivity [YWK99]. Furthermore, the Arnoldi-based model order reduction strategy defined on the Hilbert space may be utilised to yield a highly accurate reduced-order model. However, the reduction process suffers from numerical instabilities associated with explicit moment-matching. Implicit momet-matching in combination with the ICT has been recently proposed by Gad and Nakhla [GM04].

Table 3.1 summarises the most important properties of all eight of the simulation strategies for interconnect macromodeling.

	Lossless lines	Lossy lines	Frequency-domain description	Passivity	Major disadvantage	Recommend for HF interconnect
Lumped segmentation technique	YES	YES	NO	YES	Choice of section length	NO
Direct-time stepping scheme	YES	YES	NO	YES	Extremely small time-step	NO
Convolution techniques	YES	YES	NO	YES	Integration over past history	NO
The method of characteristics (MC)	YES (very suitable)	YES (generalised MC)	YES	YES	Only for lossless lines	NO
Exponential matrix-rational approx. (MRA)	YES	YES (but not well suited)	YES	NO	Not suited for long, relatively lossless lines	YES (for certain types)
Basis function approximation	YES	YES	NO	NO	Not passive in combination with MOR	YES (when passivity of interconnect model is not required)
Compact-finite-differences approximation	YES	YES	YES	YES	Complex implementa-tion	YES
Integrated congruence transform (ICT)	YES	YES	YES	YES	Numerical instability associated with explicit moment matching	YES

Table 3.1. *Strategies based on interconnect macromodeling*

3.3. Interconnect modelling based on model order reduction

A second class of interconnect modelling strategies are based on *model order reduction (MOR)* (e.g. [CN94] [SKE96], [FF95a], etc). The model order reduction strategy aims to form a good approximation of the original large interconnect system over a certain range of time and frequency, i.e. to project a larger system to the smaller one with similar behaviour. The resulting reduced order model (ROM), described with a much smaller number of state variables, may then be passed to a nonlinear simulator, e.g. SPICE, and simulated within the overall circuit as shown in Fig. 3.5. A mapping back strategy closely related to the MOR technique may be employed to determine the variables of the original model.

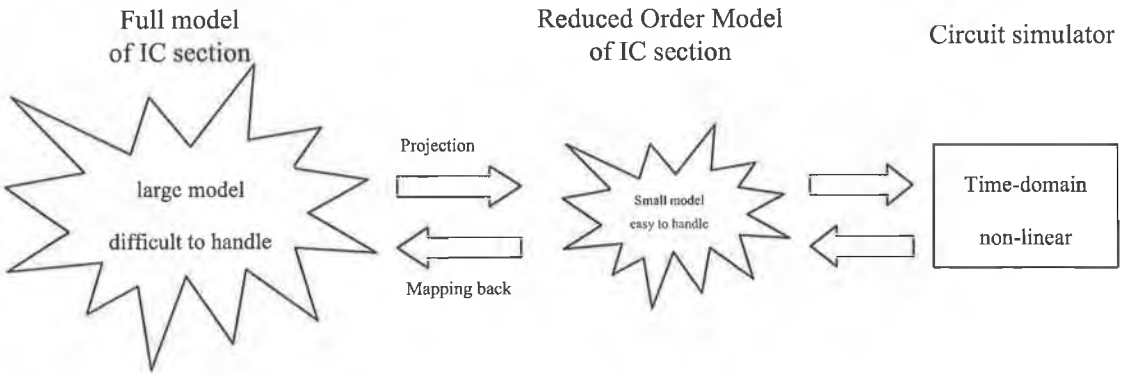


Fig. 3.5. *Reduction strategy*

There are two common ways of applying the reduction technique to an interconnect model. In the first approach, the reduction is performed during the model construction when only the important behaviour is taken into account. By introducing certain assumptions (e.g. like in PEEC method [R74]), a smaller ROM is obtained and the computational burden for a simulator is reduced. However, the price to be paid is the intrinsic inaccuracy of the overall model.

Nowadays, there is a growing demand for models that incorporate many aspects of the circuit behaviour and assumptions previously made in order to reduce the model are not justifiable anymore. This leads to the second approach in MOR where the full model incorporating all necessary parameters of a circuit is taken as a starting point. This model may be obtained from a full-wave simulator or from measurements either in the time or frequency domain. Then suitable techniques are developed to replace an initial large model by a smaller one with approximately the same behaviour as illustrated in Fig 3.5.

The research presented in this dissertation focuses on the second reduction approach since signal integrity issues in modern high-frequency interconnects require use of all of the available system parameters. The MOR algorithms may be classified into two large groups: *moment-matching based* (e.g. Asymptotic Waveform evaluation (AWE) [PR90] and Krylov subspace methods [SKE96], [FF95a]), and *singular value decomposition* (SVD) based techniques (e.g. truncated balanced realisation [M81], Hankel norm approximation [G84], etc). Gugercin and Antoulas [GA00] have shown that SVD based methods are more accurate when the whole frequency range is considered since moment matching methods always lead to higher error norms due to their local nature. But SVD-based methods are found to be extremely computationally expensive and cannot handle systems with a very high-order, e.g. large high-frequency

interconnect networks. On the other hand, moment-matching techniques, especially Krylov subspace based ones, have proved to be far superior in terms of numerical efficiency and thus appropriate for handling large systems. Therefore, from this point forward, only moment-matching MOR techniques will be considered as SVD based techniques cannot cope with the size of modern interconnect networks. In the rest of this section, a few important aspects of model order reduction schemes are discussed and some properties of several moment-matching MOR techniques are presented.

3.3.1. State space system representation

Following some initial interconnect modelling technique, the partial differential equations (PDE) that govern interconnect network behaviour are converted to a set of ordinary differential equations (ODE). Usually they are written in standard Modified Nodal Analysis (MNA) notation as [HRB75]:

$$\begin{aligned} &\boldsymbol{C}\dot{\boldsymbol{x}}(t)+\boldsymbol{G}\boldsymbol{x}(t)=\boldsymbol{B}\boldsymbol{u}(t), && \boldsymbol{C},\boldsymbol{G}\in\Re^{nxn},\ \boldsymbol{B}\in\Re^{nxl},\ \boldsymbol{x}\in\Re^{nxl} \\ &y(t)=\boldsymbol{L}^T\boldsymbol{x}(t), && \boldsymbol{L}\in\Re^{nxl} \end{aligned} \tag{3.37}$$

where n represents the total number of MNA variables. Vector $\boldsymbol{x}(t)$ is a vector of state variables (the capacitor voltages and inductor currents), $\boldsymbol{u}(t)$ is vector containing a set of inputs and $y(t))$ is vector of outputs. Matrix $\boldsymbol{C}$ represents the contribution of memory elements such as capacitor and inductors while matrix $\boldsymbol{G}$ represents that of memory-less elements such as resistors. Matrices $\boldsymbol{B}$ and $\boldsymbol{L}$ contain a description of the circuit topology and are always real constant matrices. In order to solve this ODE system the Laplace transform may be applied yielding a state space formulation as follows

$$\begin{aligned} &s\boldsymbol{C}\boldsymbol{X}(s)+\boldsymbol{G}\boldsymbol{X}(s)=\boldsymbol{B}\boldsymbol{U}(s) \\ &\boldsymbol{Y}(s)=\boldsymbol{L}^T\boldsymbol{X}(s) \end{aligned} \tag{3.38}$$

Without loss of generality, zero initial conditions ($\boldsymbol{X(0)=0}$) are assumed. The transfer function of this system in the frequency domain is defined as the ratio of the system output and system input:

$$\boldsymbol{H}(s)=\boldsymbol{U}(s)^{-1}\boldsymbol{Y}(s)=\boldsymbol{L}^T(\boldsymbol{G}+s\boldsymbol{C})^{-1}\boldsymbol{B} \tag{3.39}$$

The frequency domain function $\boldsymbol{H(s)}$ gives the full information of the system behaviour as it directly relates system inputs to the system outputs. It is independent of the value of the excitation at the input and may be used to analyse systems irrespective of input

signal. Therefore most MOR algorithms approximate a system by a reduced model that approximate the behaviour of $\boldsymbol{H}(s)$, as illustrated in Fig 3.6.

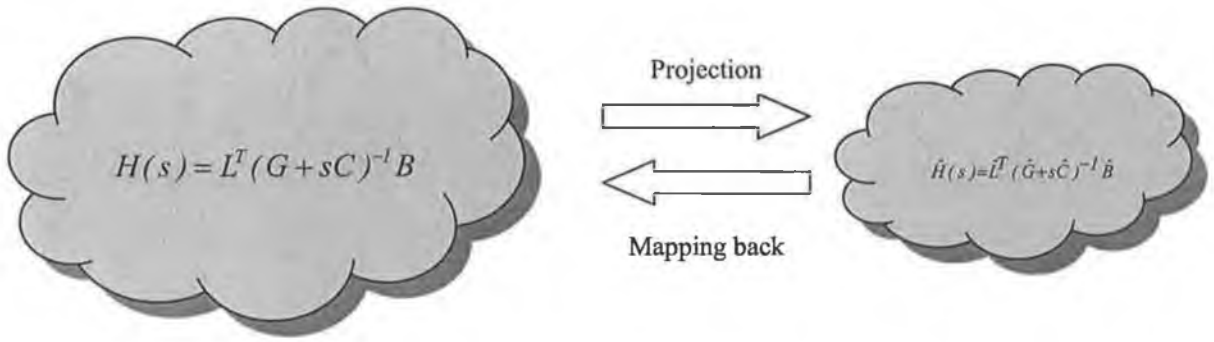


Fig. 3.6. *Model Order Reduction (MOR)*

This reduced order model described with $\hat{\boldsymbol{H}}(s)$ can then be used to approximate the time-domain or frequency-domain response of a linear circuit or interconnect over a predetermined range of excitation frequencies.

3.3.2. Rational and pole-residue system representation

The transfer function of a single input/single output system may be written in *rational form* as:

$$H(s)=\frac{P_m(s)}{Q_n(s)} \tag{3.40}$$

where $P_m(s)$ and $Q_n(s)$ are polynomials of m^{th} and n^{th} order respectively in s-domain. Alternatively, (3.40) may be written in *pole-residue representation* as:

$$H(s)=c+\sum_{i=1}^{n}\frac{k_i}{s-p_i} \tag{3.41}$$

where p_i and k_i are i^{th} pole-residue pair, constant c represents direct coupling between the system input and output and n is the total number of system poles. The time-domain representation of (3.41) is called the *impulse response* and may be analytically computed using an inverse Laplace transform as:

$$h(t)=c\delta(t)+\sum_{i=1}^{n}k_i e^{p_i t} \tag{3.42}$$

where $\delta(t)$ stands for Kronecker delta function.

In general, interconnect networks have a very large number of poles spread over a wide-frequency range. This makes simulation of such interconnect networks very CPU intensive by imposing a very small time-step on the solver in order to account for

all poles of a network. But most of the behaviour of a network is usually well characterised by a small number of, so called, *dominant poles*, i.e. poles that are close to the imaginary axis. As an example [AN01], consider a system characterised by only two poles P_1=-2 and P_2=-1000, i.e. the transfer function of such a system may be given as:

$$H(s)=\frac{1}{(s+2)(s+1000)} \tag{3.43}$$

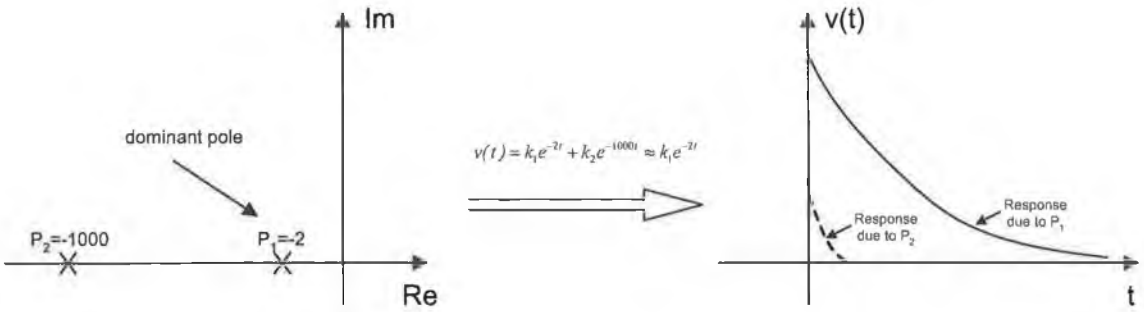


Fig. 3.7. *Dominant poles*

As may be seen, the response due to the pole P_2 (the pole that is far away from imaginary axis) is negligible after a very short time but the solver is still forced to work with the small step in order to take into account the contribution due to P_2 for the duration of the simulation.

An interconnect network will usually have a total number of poles of the order of hundreds which will be highly computationally expensive. Large networks usually have a total number of poles of the order of thousands and computing all the poles for such networks is totally impractical if not impossible. Therefore MOR techniques for the simulation of interconnect networks address this issue by deriving a reduced-order approximation $\hat{H}(s)$ in terms of q dominant poles:

$$H(s)\approx\hat{H}(s)=\frac{\hat{P}_r(s)}{\hat{Q}_q(s)}=\hat{c}+\sum_{j=1}^{q}\frac{\hat{k}_j}{s-\hat{p}_j} \tag{3.44}$$

Here p_j and k_j are the j^{th} pole-residue pair and $q<<n$ is the total number of reduced system poles. The pole-residue pairs for $\hat{H}(s)$ are determined from the condition that the q^{th} - order transfer function $\hat{H}(s)$ should match first q moments of a full order $H(s)$.

3.3.3. Matching of moments

The MOR techniques that are used for interconnect simulation are often referred to as *moment-matching techniques* due to the analogy between time-domain moments of the impulse response *h(t)* and coefficients in the Taylor-series expansion of a transfer function *H(s)* around some point in the complex plane. Consider the Taylor series expansion of a given transfer function, *H(s),* around a point $s_0=0$

$$H(s)=H(0)+\frac{(H(0))^{(1)}}{1!}s+\frac{(H(0))^{(2)}}{2!}s^2+\cdots++\frac{(H(0))^{(n)}}{n!}s^n+\cdots \tag{3.45}$$

where the superscript *(n)* denotes the n^{th} derivative of *H(s).* Denoting

$$m_i=\frac{(H(0))^{(i)}}{i!}, \tag{3.46}$$

equation (3.45) may be rewritten in a simpler notation

$$H(s)=m_0+m_1s+m_2s^2+\cdots+m_ns^n+\cdots=\sum_{i=0}^{\infty}m_is^i. \tag{3.47}$$

Approximating *H(s)* with the first *n* members of the expansion yields:

$$H(s)\approx\hat{H}(s)=m_0+m_1s+m_2s^2+\cdots+m_ns^n=\sum_{i=0}^{n}m_is^i. \tag{3.48}$$

On the other hand, using the Laplace transform of *h(t)* and the expansion of the exponential function around the point $s_0=0$, one obtains

$$\begin{aligned}H(s)&=\int_0^{\infty}h(t)e^{-st}dt=\int_0^{\infty}h(t)\left[1-st+\frac{s^2t^2}{2!}-\cdots\right]dt=\\&=\int_0^{\infty}h(t)dt+s\int_0^{\infty}(-1)h(t)dt+s^2\int_0^{\infty}\frac{t^2}{2!}h(t)dt+\cdots\end{aligned} \tag{3.49}$$

Finally, rewriting (3.49) in compact notation:

$$H(s)=\sum_{i=0}^{\infty}\left(\frac{(-1)^i}{i!}\int_0^{\infty}t^ih(t)dt\right)s^i \tag{3.50}$$

and comparing (3.46), (3.47) and (3.50) one can now write

$$m_i=\frac{(H(0))^{(i)}}{i!}=\frac{(-1)^i}{i!}\int_0^{\infty}t^ih(t)dt. \tag{3.51}$$

The relation (3.51) is the reason why m_i, the coefficients of Taylor series expansion, are often referred to as *moments*. This implies that approximating the transfer function of a network, *H(s),* in terms of dominant poles is equivalent to matching, i.e. preserving, the first *n* moments of a network.

In the general case, the transfer function *H(s)* in (3.39) may be expanded around an arbitrary point $s_0 \in \mathbb{C}$

$$H(s) = m_0 + m_1(s - s_0) + m_2(s - s_0)^2 + \cdots \tag{3.52}$$

Then a reduced order model of order q is formed with a transfer function $\hat{H}(s)$

$$\hat{H}(s) = \hat{m}_0 + \hat{m}_1(s - s_0) + \hat{m}_2(s - s_0)^2 + \cdots + \hat{m}_q(s - s_0)^q \tag{3.53}$$

such that for an appropriate q there holds

$$m_i = \hat{m}_i, \quad i = 0,1,...,q\,. \tag{3.54}$$

In order to obtain an accurate model of a network it is required that the reduced-order model preserve (or match) as many moments as possible. A simple example of matching the first moment of the response is the Elmore delay [E48], [RPH83], which essentially approximates the midpoint of the monotonic step response waveform by the mean of the impulse response [AN01].

A number of moment-matching based MOR algorithms for interconnect network simulation have been proposed in the literature [CN94], [FF95a], [CN95], [EL97]. For example, in the case when $s_0 = 0$ (the expansion in (3.52) is around the origin), the reduced-order model may be computed recursively, by means of an AWE algorithm [PR90]. In the case when $s_0 = \infty$ (i.e. when the expansion in (3.52) is around infinity), the reduced-order model may be computed by means of Arnoldi [EL97] or Lanczos procedures [FF95a]. Depending on the manner that the technique matches the moments (explicitly or implicitly), the moment-matching technique may be classified [AN01] either as an *explicit* moment-matching technique (AWE and its derivatives) or an *implicit* moment-matching technique (the techniques based on projection onto a Krylov subspace, e.g. Arnoldi, Lanczos).

3.3.4. Explicit moment-matching techniques

Explicit moment-matching techniques attempt to directly match the moments of the original system with the parameters of a new reduced-order model. Asymptotic Waveform Evaluation (AWE) and Complex Frequency Hopping (CFH) are typical representatives of this group of MOR algorithms and will be briefly described here.

3.3.4.1. Asymptotic Waveform Evaluation (AWE)

Asymptotic Waveform Evaluation (AWE) [PR90], [CN94], [TN92] uses a Padé approximant [BG81] to explicitly match moments of a Laplace domain transfer function

(3.47). Consider a transfer function $H(s)$ that is approximated by a rational function $\hat{H}(s)$ (Padé approximant) containing only a relatively small number of dominant poles (p_i):

$$H(s) \approx \hat{H}(s) = \frac{\hat{P}_r(s)}{\hat{Q}_q(s)} = \frac{a_0 + a_1 s + \cdots + a_r s^r}{1 + b_1 s + \cdots + b_q s^q} \tag{3.55}$$

where $a_0, \ldots, a_r, b_1, \ldots, b_q$ are $r+q+1$ coefficients of a Padé approximant. Matching this rational function approximation to a Taylor series expansion (3.48) in terms of moments with $n=q+r$ yields

$$\frac{a_0 + a_1 s + \cdots + a_r s^r}{1 + b_1 s + \cdots + b_q s^q} = m_0 + m_1 s + m_2 s^2 + \cdots + m_{q+r} s^{q+r}. \tag{3.56}$$

It can be shown that the Padé approximation is more accurate than the original Taylor expansion [AN04]. Cross-multiplying and equating the coefficients of s starting from s^0 and going to s^L, the coefficients of the numerator may be calculated as:

$$\begin{aligned} a_0 &= m_0 \\ a_1 &= m_1 + b_1 m_0 \\ &\vdots \\ a_r &= m_r + \sum_{i=1}^{\min(r,q)} b_i m_{r-i} \end{aligned} \tag{3.57}$$

The coefficients of the denominator polynomial are obtained in a similar manner by equating coefficients of s starting from s^{r+1} and going to s^{r+q}, yielding:

$$\begin{bmatrix} m_{r-q+1} & m_{r-q+2} & \cdots & m_r \\ m_{r-q+2} & m_{r-q+3} & \cdots & m_{r+1} \\ \vdots & \vdots & \ddots & \vdots \\ m_r & m_{r+1} & \cdots & m_{r+q-1} \end{bmatrix} \begin{bmatrix} b_q \\ b_{q-1} \\ \vdots \\ b_1 \end{bmatrix} = - \begin{bmatrix} m_{r+1} \\ m_{r+2} \\ \vdots \\ m_{r+q} \end{bmatrix}. \tag{3.58}$$

Alternatively, the AWE model may be expressed in terms of a pole-residue pair. Poles p_i are found by solving the polynomial equation:

$$\hat{Q}_q(s) = 0. \tag{3.59}$$

In order to obtain the residues k_i, the approximate transfer function is first expanded in terms of a MacLaurin series as:

$$\hat{H}(s) = \hat{c} + \sum_{i=0}^{\infty} \left(-\sum_{j=1}^{q} \frac{\hat{k}_j}{\hat{p}_j^{\,n+1}} \right) s^i. \tag{3.60}$$

Comparing (3.60) and (3.48) it can be seen that

$$\begin{aligned} m_0 &= \hat{c} - \sum_{j=1}^{r} \frac{\hat{k}_j}{\hat{p}_j} = \frac{\hat{k}_j}{\hat{p}_j} \\ &\vdots \\ m_j &= -\sum_{j=1}^{r} \frac{\hat{k}_j}{\hat{p}_j^{\,j+1}} \end{aligned} \tag{3.61}$$

When written in matrix form, (3.61) becomes

$$\begin{bmatrix} \hat{p}_1^{-1} & \hat{p}_2^{-1} & \cdots & \hat{p}_r^{-1} & -1 \\ \hat{p}_1^{-2} & \hat{p}_2^{-2} & \cdots & \hat{p}_r^{-2} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat{p}_1^{-r-2} & \hat{p}_2^{-r-2} & \cdots & \hat{p}_r^{-r-2} & 0 \\ \hat{p}_1^{-r-1} & \hat{p}_2^{-r-1} & \cdots & \hat{p}_r^{-r-1} & 0 \end{bmatrix} \begin{bmatrix} \hat{k}_1 \\ \hat{k}_2 \\ \vdots \\ \hat{k}_r \\ \hat{c} \end{bmatrix} = - \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_{r-1} \\ m_r \end{bmatrix} \tag{3.62}$$

Equations (3.57) and (3.58) or alternatively (3.59) and (3.62) give access to the coefficients of the Padé approximant or the pole-residue pair that may be calculated if moments m_i are known. It has been shown [CN94], [AN01] that it is possible to find a closed form relationship for the computation of moments. Consider the simple case of a lumped circuit described by:

$$\begin{aligned} (\boldsymbol{G}+s\boldsymbol{C})\boldsymbol{X}(s) &= \boldsymbol{b}(s) \\ \boldsymbol{y} &= \boldsymbol{L}^T\boldsymbol{X}(s) \end{aligned} \tag{3.63}$$

The Taylor series expansion of $\boldsymbol{X}(s)$ in terms of moments may be written as:

$$\boldsymbol{X}(s) = \boldsymbol{M}_0 + \boldsymbol{M}_1 s + \boldsymbol{M}_2 s^2 + \cdots \tag{3.64}$$

where $\boldsymbol{M}_i$ represents the i^{th} moment vector. Substituting (3.64) in the first equation in (3.63) yields

$$(\boldsymbol{G}+s\boldsymbol{C})(\boldsymbol{M}_0 + \boldsymbol{M}_1 s + \boldsymbol{M}_2 s^2 + \cdots) = \boldsymbol{b}\,. \tag{3.65}$$

Multiplying the left hand side and equating coefficients of the same powers of s, the following relationships are obtained

$$\left.\begin{aligned} \boldsymbol{G}\boldsymbol{M}_0 &= \boldsymbol{b} \\ \boldsymbol{G}\boldsymbol{M}_i &= -\boldsymbol{C}\boldsymbol{M}_{i-1}, \quad i>0 \end{aligned}\right\} \Rightarrow \begin{aligned} \boldsymbol{M}_0 &= \boldsymbol{G}^{-1}\boldsymbol{b} \\ \boldsymbol{M}_i &= -\boldsymbol{G}^{-1}\boldsymbol{C}\boldsymbol{M}_{i-1}, \quad i>0 \end{aligned} \tag{3.66}$$

The moments necessary to calculate the Padé coefficients in (3.57) and (3.58) or, alternatively, the poles and residues in (3.59) and (3.62) are taken from moment-vectors $\boldsymbol{M}_i$. The cost to calculate the moments of a single-input single-output system is one LU decomposition. Therefore, AWE provides a significant computational speed up when compared to the conventional SPICE algorithm (up to 1000 times faster) [AN01]. In the case of networks containing distributed lines, moment computation is not

straightforward but it can be done [TN92], [YK95], [AN01]. However, the number of dominant poles will be significantly higher and a single-point Padé expansion is often unable to capture all of them.

AWE tends not to be used in modern simulators due to its serious limitations. It stagnates in accuracy when the order of the approximation increases. The moment-matrix in (3.58) is extremely ill-conditioned. Furthermore, AWE often produces unstable poles in the reduced system. Accuracy deteriorates when far from expansion point as AWE is only capable of capturing poles around the origin as illustrated in Fig. 3.8. It does not provide estimates for error bounds and it does not guarantee passivity [AN01]. Some of the limitations of AWE may be overcome using a multipoint expansion technique such as Complex Frequency Hopping.

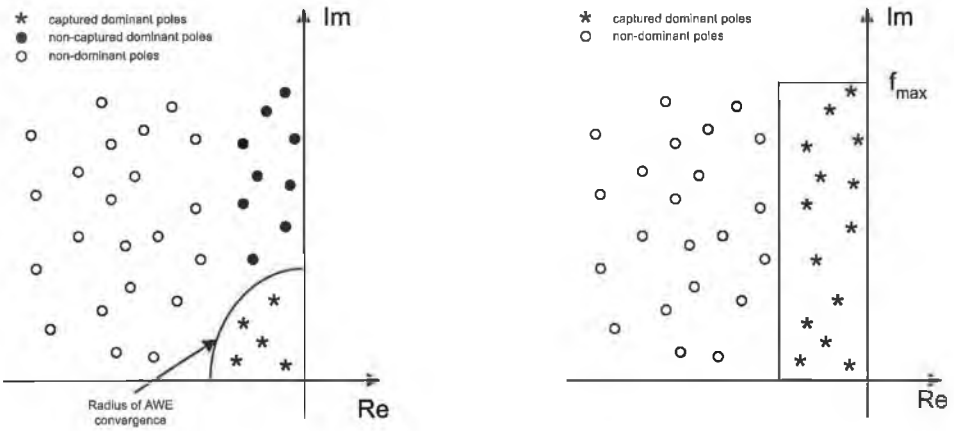
3.3.4.2. Complex Frequency Hopping (CFH)

Complex Frequency Hopping (CFH) [CN95], [AN01] extends the process of explicit moment matching to multiple expansion points, called *hops*, in the complex plane near or on the imaginary axis up to a predefined highest frequency of interest. CFH relies on a *binary search algorithm* to determine the expansion points and to minimise the number of expansions.

In the case of expansion at an arbitrary point, the moments may be calculated in a similar manner to (3.66):

$$\left.\begin{aligned}&(\boldsymbol{G}+s_0\boldsymbol{C})\boldsymbol{M}_0=\boldsymbol{b}\\&(\boldsymbol{G}+s_0\boldsymbol{C})\boldsymbol{M}_i=-\boldsymbol{C}\boldsymbol{M}_{i-1},\quad i>0\end{aligned}\right\}\Rightarrow\quad\begin{aligned}&\boldsymbol{M}_0=(\boldsymbol{G}+s_0\boldsymbol{C})^{-1}\boldsymbol{b}\\&\boldsymbol{M}_i=-(\boldsymbol{G}+s_0\boldsymbol{C})^{-1}\boldsymbol{C}\boldsymbol{M}_{i-1},\quad i>0\end{aligned}\tag{3.67}$$

Using the information from all the expansion points, CFH extracts a dominant pole set as illustrated [AN01] in Fig 3.8.



a) Dominant poles from AWE *b) Dominant poles from CFH*

Fig. 3.8. *AWE and CFH dominant poles*

A Padé approximation is accurate only near the point of expansion. Moving away from the expansion point, the accuracy of the approximation decreases and in order to validate it, at least two expansion points are necessary. The accuracies of these two expansions can be verified either by a *pole-matching-based approach* (matching poles generated at both hops) [CN95] or a *transfer-function-based approach* (comparing the value of the transfer functions produced by both hops at a point intermediate to them) [SCN+94].

CFH produces poles that are guaranteed to be stable up to a user defined frequency point. Although CFH provides an error criterion for the selection of accurate poles, it still suffers from an ill-conditioning problem. Furthermore, passivity is not guaranteed [AN01].

3.3.4.3. Some comments on ill-conditioning

Consider the time-domain MNA equations given by (3.37). Multiplying (3.37) with $\boldsymbol{G}^{-1}$ one can write

$$\begin{aligned} \boldsymbol{A}\dot{\boldsymbol{x}}(t) &= \boldsymbol{x}(t) - \boldsymbol{b}u(t) \\ y(t) &= \boldsymbol{L}^T \boldsymbol{x}(t) \end{aligned} \tag{3.68}$$

where $\boldsymbol{A} = -\boldsymbol{G}^{-1}\boldsymbol{C}$, $\boldsymbol{b} = \boldsymbol{G}^{-1}\boldsymbol{B}$. Taking the Laplace transform of (3.68), one can write the equations in the frequency domain:

$$\begin{aligned} s\boldsymbol{A}\boldsymbol{X}(s) &= \boldsymbol{X}(s) - \boldsymbol{b}U(s) \\ Y(s) &= \boldsymbol{L}^T \boldsymbol{X}(s) \end{aligned} \tag{3.69}$$

The transfer function H_{sys} of a given system is now written as:

$$H_{sys}(s) = \frac{Y(s)}{U(s)} = \boldsymbol{L}^T (\boldsymbol{I} - s\boldsymbol{A})^{-1} \boldsymbol{b}, \tag{3.70}$$

where $\boldsymbol{I}$ is identity matrix of dimension n. Expanding the middle term in terms of a Taylor series, one can write

$$H_{sys}(s) = \boldsymbol{L}^T (\boldsymbol{I} + s\boldsymbol{A} + s^2\boldsymbol{A}^2 + \cdots s^q \boldsymbol{A}^q)\boldsymbol{b} = \sum_{k=0}^{q} (\boldsymbol{L}^T \boldsymbol{A}^k \boldsymbol{b}) s^k \tag{3.71}$$

Comparing (3.71) to (3.48) one can write the moments as

$$m_k = \boldsymbol{L}^T \boldsymbol{A}^k \boldsymbol{b}. \tag{3.72}$$

As can be seen, when successive moments are explicitly calculated, they are obtained in terms of powers of $\boldsymbol{A}$. As k increases (which corresponds to obtaining higher-order moments), the process quickly converges to the eigenvector corresponding to the

eigenvalue of $\boldsymbol{A}$ with the largest magnitude [AN01]. As a result, for relatively large values of k, the explicitly calculated moments m_k, m_{k+1}, m_{k+2}, ..., do not add any extra information to the moment matrix as all of them contain information only about the largest eigenvalue making the rows beyond k of the moment-matrix almost identical. This is the reason why increasing the order of the Padé approximation (which is equivalent to matching more moments) does not give a better approximation. Moreover, it results in a moment matrix that is extremely ill-conditioned [AN01].

In order to overcome the two major drawbacks of the explicit moment-matching techniques, the ill-conditioning of the moment matrix and the non-preservation of passivity, indirect moment-matching techniques have been developed. These techniques are based on the *Krylov subspace formulation* and *congruent transformation* and very often are referred to as *Krylov techniques*.

3.3.5. Implicit moment-matching techniques (Krylov techniques)

Unlike explicit moment matching techniques (AWE and CFH) which form a reduced model based on extracting the *dominant poles* of a given system, implicit moment-matching techniques aim to construct a reduced model based on the extraction of the *leading eigenvalues* (eigenvalues with the largest magnitude) of a given system [AN01].

Consider (3.68) and assume that the matrix $\boldsymbol{A}$ can be diagonalized in the form

$$\boldsymbol{A} = \boldsymbol{F}\boldsymbol{\lambda}\boldsymbol{F}^{-1}, \tag{3.73}$$

where $\boldsymbol{\lambda} = diag\,[\lambda_1\ \lambda_2\ \cdots\ \lambda_n]$ is a diagonal matrix containing eigenvalues of matrix $\boldsymbol{A}$ and matrix $\boldsymbol{F}$ contains the eigenvectors of matrix $\boldsymbol{A}$. The transfer function may now be written as:

$$H_{sys}(s) = \boldsymbol{L}^T(\boldsymbol{I} - s\boldsymbol{F}\boldsymbol{\lambda}\boldsymbol{F}^{-1})^{-1}\boldsymbol{b} = \boldsymbol{L}^T\boldsymbol{F}(\boldsymbol{I} - s\boldsymbol{\lambda})^{-1}\boldsymbol{F}^{-1}\boldsymbol{b} = \boldsymbol{L}^T\boldsymbol{F}\begin{bmatrix} \frac{1}{1-s\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{1-s\lambda_n} \end{bmatrix}\boldsymbol{F}^{-1}\boldsymbol{b} \tag{3.74}$$

Equation (3.74) may be written as:

$$H_{sys}(s) = \sum_i \frac{\eta_i}{1-s\lambda_i} = \sum_i \frac{-\eta_i/\lambda_i}{s-1/\lambda_i} = \sum_i \frac{k_i}{s-p_i} \tag{3.75}$$

where η_i are functions relating to $\boldsymbol{L}$, $\boldsymbol{F}$ and $\boldsymbol{b}$. Upon close inspection of (3.75) one can draw the conclusion that the poles p_i are the reciprocal of the eigenvalues λ_i of the matrix $\boldsymbol{A}$ [AN01]. The leading eigenvalues, i.e. the eigenvalues with the largest magnitudes, correspond to the poles closer to the origin. If the eigenvalues and eigenvectors of $\boldsymbol{A}$ are obtained, the transfer function in terms of poles and residues may easily be obtained.

Large interconnect networks are characterised by a great number of eigenvalues and eigenvectors and it would be highly impractical if not impossible to calculate all of them. Therefore reduction techniques that extract the leading eigenvalues using projection to the Krylov subspace were developed.

3.3.5.1. Krylov subspace method

Consider the circuit equations (3.68) and a simple similarity transform (Appendix A):

$$\boldsymbol{AK} = \boldsymbol{KH}_n, \tag{3.76}$$

where $\boldsymbol{K}$ is the transformation matrix defined as:

$$\boldsymbol{K} = [\boldsymbol{b}\ \boldsymbol{Ab} \cdots \boldsymbol{A}^{n-1}\boldsymbol{b}] \tag{3.77}$$

and $\boldsymbol{H}_n$ is the upper-Hessenberg matrix of dimension n (Appendix A). Since $\boldsymbol{H}_n$ is related to the matrix $\boldsymbol{A}$ through a similarity transformation, its eigenvalues are the same as that of $\boldsymbol{A}$. However, direct computation of $\boldsymbol{H}_n$ has a couple of limitations. Computing $\boldsymbol{H}_n$ as

$$\boldsymbol{H}_n = \boldsymbol{K}^{-1}\boldsymbol{AK} \tag{3.78}$$

requires the inverse of the dense matrix $\boldsymbol{K}$ and hence, its computation is very expensive. Also, $\boldsymbol{K}$ is likely to be ill-conditioned as it is formed based on the sequence, $\boldsymbol{A}^i\boldsymbol{R}$, which quickly converges to the eigenvector corresponding to the largest eigenvalue. Thus it has the same problem as with explicit moment-matching techniques.

To overcome these problems, it has been suggested to replace the matrix $\boldsymbol{K}$ with the orthogonal matrix $\boldsymbol{Q}$ such that for all n, the leading n columns of $\boldsymbol{K}$ and $\boldsymbol{Q}$ *span the same space* that is called the Krylov subspace $\mathcal{K}_n(\boldsymbol{A},\boldsymbol{b})$ and noted as:

$$\mathcal{K}_n(\boldsymbol{A},\boldsymbol{b}) = span([\boldsymbol{b}\ \boldsymbol{Ab} \cdots \boldsymbol{A}^{n-1}\boldsymbol{b}]) = span([\boldsymbol{Q}]) \tag{3.79}$$

Mathematically, it means that any vector that is a linear combination of the leading n columns of $\boldsymbol{K}$ can be expressed as linear combination of the leading n columns of $\boldsymbol{Q}$. In

contrast to $\boldsymbol{K}$, the orthogonal matrix $\boldsymbol{Q}$ is well conditioned and easily invertible since $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{T}$. Therefore, expressing matrix $\boldsymbol{K}$ as:

$$\boldsymbol{K} = \boldsymbol{Q}\boldsymbol{R}_u \tag{3.80}$$

where $\boldsymbol{R}_u$ is an upper-triangular matrix and substituting in (3.78), yields

$$\boldsymbol{H}_n = \boldsymbol{K}^{-1}\boldsymbol{A}\boldsymbol{K} = (\boldsymbol{Q}\boldsymbol{R}_u)^{-1}\boldsymbol{A}(\boldsymbol{Q}\boldsymbol{R}_u) = (\boldsymbol{R}_u^{-1}\boldsymbol{Q}^{T})\boldsymbol{A}(\boldsymbol{Q}\boldsymbol{R}_u). \tag{3.81}$$

Multiplying (3.81) with $\boldsymbol{R}_u$ on the left hand side and $\boldsymbol{R}_u^{-1}$ on the right hand side yields:

$$\boldsymbol{R}_u\boldsymbol{H}_n\boldsymbol{R}_u^{-1} = \boldsymbol{Q}^{T}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{H} \tag{3.82}$$

Matrix $\boldsymbol{H}$ is also in upper Hessenberg form since $\boldsymbol{R}_u$ and $\boldsymbol{R}_u^{-1}$ are upper triangular and $\boldsymbol{H}_n$ is an upper Hessenberg matrix (Appendix A).

If now only the leading q columns $(q<n)$ of $\boldsymbol{Q}$ are used, the dimension of the matrix $\boldsymbol{Q}$ will be nxq, yielding $\boldsymbol{H} \rightarrow \boldsymbol{H}_q \in \Re^{q\times q}$. This means that using an orthogonal transformation, matrix $\boldsymbol{A}$ of dimension nxn is reduced to a smaller upper Hessenberg matrix $\boldsymbol{H}_q$ of dimension qxq. Another very important property is that the columns of $\boldsymbol{Q} = [\boldsymbol{q}_0, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_{q-1}] \in \Re^{n\times q}$, $\boldsymbol{q}_i$ - orthogonal vectors, can be computed one at a time giving the benefit of computing only the columns of $\boldsymbol{Q}$ that are needed [AN01].

Recently, several techniques for the simulation of interconnect networks based on Krylov subspace projections have been developed, most notably PRIMA (based on the Arnoldi algorithm) and Padé Via Lanczos (PVL).

3.3.5.2. MOR based on the Arnoldi process

Consider the Krylov space

$$\mathcal{K}_q(\boldsymbol{b}, \boldsymbol{A}) = span[\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, \ldots, \boldsymbol{A}^{q-1}\boldsymbol{b}] = span([\boldsymbol{Q}]) \tag{3.83}$$

To implement the Arnoldi algorithm for circuit order reduction, the vector $\boldsymbol{x}$ of dimension n is mapped into a smaller vector $\hat{\boldsymbol{x}}$ of dimension q $(q<<n)$ using a congruent transformation:

$$\boldsymbol{x}_{n\times 1} = \boldsymbol{Q}_{n\times q}\,\hat{\boldsymbol{x}}_{q\times 1}, \tag{3.84}$$

where $\boldsymbol{Q}$ is orthogonal matrix. In that case, the transfer function $H_{sys}(s)$ is written as:

$$\boldsymbol{H}_{sys}(s) = \boldsymbol{L}^{T}(\boldsymbol{G}+s\boldsymbol{C})^{-1}\boldsymbol{B} = \boldsymbol{L}^{T}(\boldsymbol{I}-s\boldsymbol{A})^{-1}\boldsymbol{R}, \tag{3.85}$$

where $\boldsymbol{A} = -\boldsymbol{G}^{-1}\boldsymbol{C}$ and $\boldsymbol{R} = \boldsymbol{G}^{-1}\boldsymbol{B}$ maps into

$$\hat{\boldsymbol{H}}_{sys}(s) = \boldsymbol{L}^{T}\boldsymbol{Q}(\boldsymbol{I}-s\boldsymbol{Q}^{T}\boldsymbol{A}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{T}\boldsymbol{R} = \boldsymbol{L}^{T}\boldsymbol{Q}(\boldsymbol{I}-s\boldsymbol{H}_q)^{-1}\boldsymbol{Q}^{T}\boldsymbol{R}, \tag{3.86}$$

where $\boldsymbol{H}_q$ is Hessenberg matrix of dimension q. In this case the ROM may be noted as:

$$\begin{aligned} \hat{\boldsymbol{A}}\hat{\boldsymbol{x}}(t) &= \hat{\boldsymbol{x}}(t) - \hat{\boldsymbol{b}}u(t) \\ y(t) &= \hat{\boldsymbol{L}}^T \hat{\boldsymbol{x}}(t) \end{aligned}, \tag{3.87}$$

where

$$\hat{\boldsymbol{A}} = \boldsymbol{Q}^T \boldsymbol{A}\boldsymbol{Q} = \boldsymbol{H}_q, \quad \hat{\boldsymbol{b}} = \boldsymbol{Q}^T \boldsymbol{b} \text{ and } \hat{\boldsymbol{L}}^T = \boldsymbol{L}^T \boldsymbol{Q}. \tag{3.88}$$

As can be seen, the Arnoldi algorithm reduces $\boldsymbol{A}$ to a small block upper Hessenberg matrix $\boldsymbol{H}_q$. The eigenvalues of $\hat{\boldsymbol{H}}_{sys}(s)$ are given by the eigenvalues of $\boldsymbol{H}_q$ that are a good approximation to the leading eigenvalues of $\boldsymbol{A}$. Therefore, the eigenvalues of the transfer function of the reduced system (3.86) are a good approximation to the poles of the original transfer function (3.85).

Although the moments of the MNA equations (3.68) are matched during the Arnoldi process, there is no need to explicitly compute the product $\boldsymbol{A}^{q-1}\boldsymbol{b}$. Hence the ill-conditioning problem arising due to the quick convergence of the sequence $[\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, ..., \boldsymbol{A}^{q-1}\boldsymbol{b}]$ to the eigenvector of the largest eigenvalue is avoided. If q is chosen such as $q << n$, i.e. the number of columns in the Krylov-space is much smaller than the number of columns of the system matrices, the size of resultant system is reduced.

One widely used implementation of the Arnoldi process is PRIMA (Passive Reduced-order Interconnect Macromodelling Algorithm) [OCP98]. PRIMA extends the block Arnoldi process to guarantee passivity. The basic Arnoldi algorithm starts with a circuit description in the form of (3.68) and then performs a *congruent transform* as illustrated in Fig 3.9. yielding a ROM whose passivity is not guaranteed.

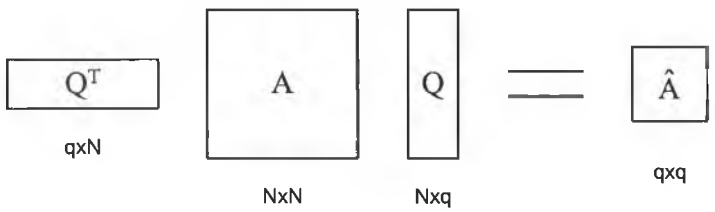


Fig. 3.9. *Congruent transformation (Arnoldi process)*

However, PRIMA starts with a system description in the form of (3.37) and performs a *split congruent transformation* where both the $\boldsymbol{G}$ and $\boldsymbol{C}$ matrices are transformed as illustrated in Fig 3.10.

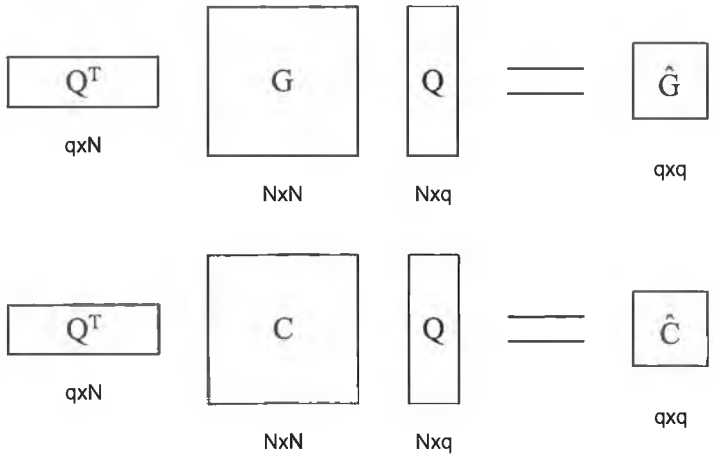


Fig. 3.10. *Split congruent transformation (PRIMA)*

It can be proven [OCP98] that if $\boldsymbol{G}$ and $\boldsymbol{C}$ are symmetric nonnegative matrices, then the reduced model

$$\begin{aligned} \hat{\boldsymbol{C}}\dot{\hat{\boldsymbol{x}}}(t)+\hat{\boldsymbol{G}}\hat{\boldsymbol{x}}(t)&=\hat{\boldsymbol{B}}\boldsymbol{u}(t) \\ \hat{\boldsymbol{y}}(t)&=\hat{\boldsymbol{L}}^T\hat{\boldsymbol{x}}(t) \end{aligned} \tag{3.89}$$

where

$$\boldsymbol{x}=\boldsymbol{Q}\hat{\boldsymbol{x}},\ \hat{\boldsymbol{C}}=\boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q},\ \hat{\boldsymbol{G}}=\boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q},\ \hat{\boldsymbol{B}}=\boldsymbol{Q}^T\boldsymbol{B} \text{ and } \hat{\boldsymbol{L}}^T=\boldsymbol{L}^T\boldsymbol{Q}, \tag{3.90}$$

is in fact passive. Consequently, in PRIMA, the transfer function is approximated by

$$\hat{\boldsymbol{H}}(s)=\boldsymbol{L}^T\boldsymbol{Q}(s\boldsymbol{Q}^T\boldsymbol{C}\boldsymbol{Q}+\boldsymbol{Q}^T\boldsymbol{G}\boldsymbol{Q})^{-1}\boldsymbol{Q}^T\boldsymbol{B}. \tag{3.91}$$

The Arnoldi method is often referred to as the Block Arnoldi method [EL97] because it can work with several columns at the same time thereby enabling straightforward implementation for multiport systems. However, such an implementation is not always efficient. It can be proven that the reduced system of order q obtained via the Arnoldi algorithm preserves the first q moments of the original network [OCP98]. However, the Padé approximant of order q in the explicit moment-matching AWE algorithm matches the first $2q$ moments. Therefore, the reduced model from the Arnoldi method will have double the size of the reduced model obtained from a direct Padé based approximation for a comparable accuracy. On the other hand, since the ill-conditioning associated with direct moment-matching algorithms is avoided, the accuracy of the Arnoldi approximation gradually increases as the order q increases, which is not the case with the Padé approximant. This redundancy in the Arnoldi algorithm is overcome with a method based on the Lanczos process. This is termed the Padé Via Lanczos (PVL) process and it preserves the first $2q$ moments [FF95a] of a network.

3.3.5.3. Padé Via Lanczos (PVL)

Padé Via Lanczos (PVL) [FF95a] is an efficient and robust implementation of the Padé approximation strategy, requiring the same amount of computations as AWE but not suffering from the problem of ill-conditioning if the order of approximation is increased. In addition, PVL is more accurate than the Arnoldi based algorithm for the same order of approximation as it matches double the number of moments.

PVL generates two bi-orthogonal Krylov-spaces:

$$\begin{aligned} \mathcal{K}_q(\boldsymbol{b}, \boldsymbol{A}) &= span[\boldsymbol{b}, \boldsymbol{A}\boldsymbol{b}, ..., \boldsymbol{A}^{q-1}\boldsymbol{b}] = span([\boldsymbol{Q}]) \\ \mathcal{K}_q(\boldsymbol{b}, \boldsymbol{A}^T) &= span[\boldsymbol{b}, \boldsymbol{A}^T\boldsymbol{b}, ..., (\boldsymbol{A}^T)^{q-1}\boldsymbol{b}] = span([\boldsymbol{P}]) \end{aligned}, \tag{3.92}$$

where $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{T}_q$ and $\boldsymbol{P}^T\boldsymbol{Q} = \boldsymbol{I}$. In a manner similar to the Arnoldi method, the congruent transformation:

$$\boldsymbol{x}_{n\times 1} = \boldsymbol{P}_{n\times q}\, \hat{\boldsymbol{x}}_{q\times 1} \tag{3.93}$$

is used and the following reduced order model is formed:

$$\begin{aligned} \hat{\boldsymbol{A}}\dot{\hat{\boldsymbol{x}}}(t) &= \hat{\boldsymbol{x}}(t) - \hat{\boldsymbol{b}}u(t) \\ y(t) &= \hat{\boldsymbol{L}}^T\hat{\boldsymbol{x}}(t) \end{aligned}, \tag{3.94}$$

where

$$\hat{\boldsymbol{A}} = \boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{P} = \boldsymbol{T}_q,\ \hat{\boldsymbol{b}} = \boldsymbol{Q}^T\boldsymbol{b} \text{ and } \hat{\boldsymbol{L}}^T = \boldsymbol{L}^T\boldsymbol{P}. \tag{3.95}$$

By running q steps of the Lanczos algorithm the matrix $\boldsymbol{A}$ is transformed to the tridiagonal matrix $\boldsymbol{T}_q$ that is the projection of the matrix $\boldsymbol{A}$ onto $\mathcal{K}_q(\boldsymbol{b}, \boldsymbol{A})$ and orthogonal to $\mathcal{K}_q(\boldsymbol{b}, \boldsymbol{A}^T)$.

The PVL algorithm is fast and accurate but its implementation for multiport systems is very complex, e.g. Matrix PVL (MPVL) [FF95b]. The disadvantage of the Lanczos algorithm is that it does not guarantee passivity of the model by construction.

3.3.6. Simulation issues related to MOR techniques

Model order reduction techniques have been acknowledged as an indispensable CAD tool that enable simulation of complex interconnect structures with reasonable accuracy. However, issues such as model stability, ill-conditioning of large matrices and ensuring passivity of the reduced model need to be addressed.

3.3.6.1. Stability

Obviously, the first condition that any reduced order model needs to fulfil is to approximate the external behaviour of the original system with sufficient engineering accuracy. However, finding a good approximation does not guarantee *stability* of the model. Clearly, a model that is not stable will introduce unphysical behaviour into the simulation and a circuit simulator will fail to converge. Therefore, any reduction strategy has to guarantee stability of the resulting ROM.

In the time domain, a system is said to be stable if a signal stays bounded for all inputs, i.e.

$$|y(x(t))| < \infty, \quad \forall x, t \in \Re . \tag{3.96}$$

In the frequency domain, stability of a system is defined in terms of poles, i.e. a linear system is said to be stable if the poles of its associated transfer function *H(s)* all have nonpositive real parts [H01]. For example, consider the transfer function *H(s)* of a circuit described by the MNA representation given in (3.39). Since matrices $\boldsymbol{L}$ and $\boldsymbol{B}$ are always real and constant matrices, it is clear that poles occur for all s for which $(\boldsymbol{G}+s\boldsymbol{C})$ in (3.39) is singular with the resultant poles of the system equal to the inverse of the eigenvalues of $-\boldsymbol{G}^{-1}\boldsymbol{C}$. Therefore, the condition of system stability may be expressed as

$$\boldsymbol{x}^{*T}(\boldsymbol{G}+s\boldsymbol{C})\boldsymbol{x}^{T} \le 0, \tag{3.97}$$

i.e. matrix $(\boldsymbol{G}+s\boldsymbol{C})$ should be negative semidefinite.

Regarding projection algorithms, it can be proven [H01] that if the projection matrix $\boldsymbol{Q}$ is a real matrix, stability will be preserved. Although it is always possible to obtain an asymptotically stable model by simply discarding the unstable poles, passivity of such a model cannot be guaranteed [OCP98].

3.3.6.2. Ill-conditioning of large matrices

Model order reduction based techniques often suffer from ill-conditioning of the large matrices that are used during the reduction process (e.g. moment matrix $\boldsymbol{M}$ in AWE or Krylov subspace matrix $\boldsymbol{K}$). Especially vulnerable are explicit moment-matching techniques, since explicit calculation of higher-order moments does not add any extra information to the moment-matrix as would be expected. Instead, the rows corresponding to higher-order moments in the moment-matrix become almost identical leading to numerical instability. The problem of ill-conditioning is partly circumvented by using projection onto a Krylov subspace to implicitly match moments without the

need to directly calculate them. However, for very large systems ill-conditioning of the Krylov subspace matrix $\boldsymbol{K}$ also arises [AN04].

3.3.6.3. Passivity

The physical nature of interconnect network is such that it cannot generate energy. In systems theory, a system is said to be passive if it cannot generate more energy that it absorbs. Hence, an interconnect network is a passive system and the reduced order model approximating interconnect behaviour should reflect this property. Unfortunately, only very few interconnect modelling techniques produce guaranteed passive interconnect models, e.g. direct lumped segmentation, integrated congruence transform, compact finite difference technique and exponential matrix-rational approximation. The situation is made worse when a reduction process is applied as most reduction techniques do not guarantee preservation of the passivity of the original model (the exception being the PRIMA algorithm that is passive by construction).

The issue of interconnect model passivity becomes very important if the reduced interconnect model is to be interfaced with a circuit simulator (e.g. SPICE). It is well known in systems theory that connecting two stable systems does not necessarily result in a stable system [RN81]. On the other hand, strictly passive circuits are asymptotically stable and arbitrary interconnections of strictly passive circuits are strictly passive [RN81]. Hence, interconnection of passive circuits will result in stable systems [OC97]. This situation is illustrated in Fig 3.11. Consequently, if an interconnect model is to be connected with other functional blocks, an additional condition of macromodel passivity must be fulfilled.

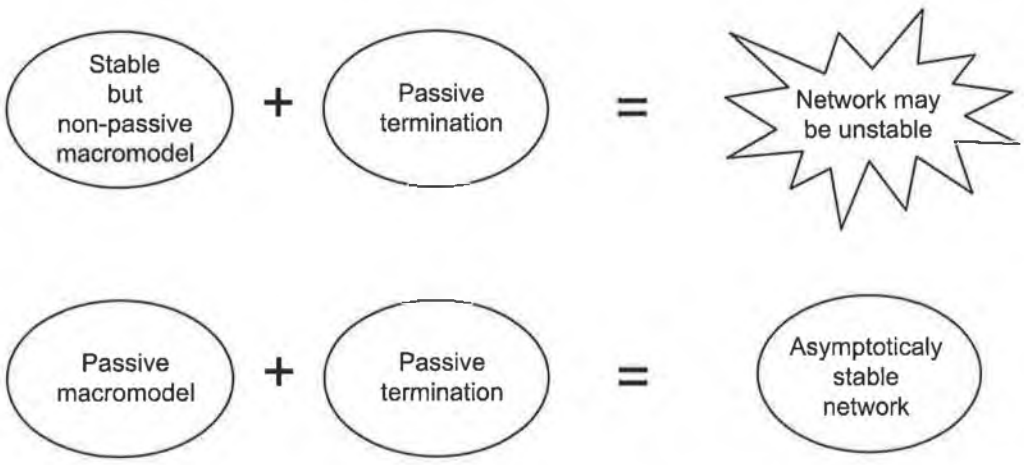


Fig. 3.11. *Passivity issue*

A system is said to be *passive* if its transfer function $\boldsymbol{H}(s)$ is positive real, i.e. it satisfies following conditions [CC98], [OCP98]:

a) Each element of $\boldsymbol{H}(s)$ is analytic for $Re\{s\} > 0$

b) $\boldsymbol{H}(s^*) = \boldsymbol{H}^*(s)$ for all complex s

c) $z^{*T}[(\boldsymbol{H}^*)^T(s) + \boldsymbol{H}(s)]z \geq 0$, for any complex vector z and all s such as $Re\{s\} > 0$.

The first condition is not necessary in the case of reduction techniques based on a real projection matrix [OCP98], since in such a case, the third condition already implies the analyticity of $\boldsymbol{H}(s)$. The second condition simply means that coefficients should be real numbers. This condition is automatically satisfied for the MNA representation reduced via a real projection matrix [OCP98]. The third condition states that $\boldsymbol{H}$(s) should be positive real matrix and in general is very difficult to prove.

3.4. Summary

The nature of modern high-frequency interconnect systems is such that utilisation of sophisticated interconnect models based on the Telegrapher's Equations description of the line behaviour is necessary. A brief overview of modern strategies for interconnect network modelling have been presented in this chapter. They are broadly classified into two main categories: *approaches based on macromodelling* (the lumped segmentation technique, the direct time-stepping scheme, convolution techniques, the method of characteristics (MC), the exponential matrix-rational approximation (EMRA), the basis function approximation, the compact-finite-differences approximation and the integrated congruence transform (ICT)) and approaches based on *model order reduction (MOR)* such as explicit moment-matching techniques (asymptotic waveform evaluation (AWE) and complex frequency hopping (CFH)) and implicit moment-matching Krylov techniques (the Arnoldi algorithm and the Lanczos process).

Macromodelling strategies aim to obtain an interconnect description in the form of a set of ordinary differential equations called the macromodel. This is usually achieved by introducing some kind of discretization of the Telegrapher's Equations. Direct discretization of the Telegrapher's Equations results in a lumped model that is highly computationally involved and inadequate to describe distortions due to high-frequency effects. The direct time-stepping scheme and convolution approaches are

extremely computationally expensive and hence cannot be considered as serious candidates for practical implementation. The method of characteristics (MC) is suitable for long low-loss lines while the exponential matrix-rational approximation (EMRA) provides a more accurate and faster model for shorter lines. For non-uniform interconnects good simulation results are achieved either via techniques that approximate the variations in space of interconnect voltages and currents in the time-domain (basis function approximation) or in the frequency domain (compact-finite-difference approximation). Utilising the integrated congruence transform (ICT), each distributed line is modelled by a lower-order passive model that has explicitly matched moments.

The aim of a *model order reduction* technique is to replace the large circuit model with a smaller one that has approximately the same behaviour. As a result, the computational burden is reduced and simulation time is shortened. Asymptotic waveform evaluation (AWE) and complex frequency hopping (CFH) techniques, based on moment matching and Padé approximation, were the first model order reduction techniques to be used in circuit analysis. However, they have intrinsic problems regarding numerical stability. The appearance of circuit simulators based on the Krylov subspace projection (the Lanczos process and the Arnoldi algorithm) solved the problem of numerical instability. The introduction of MOR algorithms for *large interconnect network* simulation has brought new issues that require developer attention. Numerical problems regarding ill-conditioning of large system matrices may lead to failure to produce a required result. Therefore, the reduction strategies have to be carefully chosen in order to minimise the possibility of numerical instabilities. The need for ensuring stability of an overall circuit consisting of several interconnected blocks highlighted the need for passive macromodels and preservation of passivity during the reduction process. Currently, there are only few techniques that are successful in preserving passivity, e.g. PRIMA.

Krylov techniques have been widely used due to their general applicability and small computational requirements when compared to other simulation techniques. They are accurate techniques and both stable and passive implementations have been developed. In particular, the Lanczos-based strategy is effective as it offers the numerical stability that AWE and CFH lack and better accuracy when compared to Arnoldi algorithm.

CHAPTER 4

Development of Interconnect Models from the Telegrapher's Equations

A novel, highly efficient and accurate technique for modelling non-uniform interconnects with frequency-dependant parameters is presented in this chapter. The technique is based on a novel form of resonant analysis theory that solves the Telegrapher's Equations in terms of the natural modes of oscillation. The resonant analysis was recently introduced by Condon [C98] to model transformers. A highly accurate resonant model is initially formed in the frequency domain to enable the frequency-dependant parameters of an interconnect to be taken into account. Contrary to the conventional approach presented in Section 3.1.4, an *a priori* assumption of line uniformity is not necessary for model derivation. Thus, the model can readily be applied to both uniform and non-uniform lines.

The contribution presented in this chapter greatly improves the efficiency of the resonant model by utilising two model order reduction approaches. The first model order reduction strategy is based on the specific structure of the resonant model. The reduced order model is obtained by neglecting the higher modes of oscillation that correspond to frequencies beyond the maximum frequency of interest. This maximum frequency of interest defines the model bandwidth and may be explicitly identified. On the other hand, the internal structure of the model is such that it allows straightforward conversion from the representation in the frequency domain to a time domain state space representation with minimal loss in accuracy. This has two distinctive advantages. First, the Lanczos-based model order reduction scheme may readily be applied to further improve the efficiency of the proposed interconnect modelling technique. Secondly, such a time-domain representation of an interconnect has the potential for straightforward incorporation into a SPICE-like circuit simulator. This addresses the problem of mixed time/frequency domain simulation described in Section 2.4.1 in an efficient and accurate manner.

4.1. Resonant analysis

Resonant analysis was recently developed as a method for modelling transmission lines [WC98] based on the application of the resonant theory developed for transformers [C98]. Initially, a model prototype is formed in the frequency domain in order to take into account the frequency-dependent parameters of an interconnect. The traditional approach to forming a frequency-domain model is based on a travelling-wave structure (3.14) obtained from conventional solution of the Telegrapher's Equations (3.3). In contrast to this, the resonant model is based on a completely new approach to the solution of the Telegrapher's Equations, in which the equations are solved in terms of the natural modes of oscillation on the transmission line. The method necessarily gives the same results as the travelling-wave approach in the frequency domain if the line is represented using exact-π sections.

4.1.1. Introduction

In the traditional approach to solving the Telegrapher's Equations, the assumption of a longitudinally *uniform* transmission line is necessary to proceed with *differentiation* of (3.6) and (3.7) to obtain decoupled equations for voltage (3.9) and current (3.10). The solution to these equations may then be written in the well-known form of travelling waves (3.14). However, as illustrated in Chapter 2, the assumption of longitudinal uniformity is not justifiable when modelling modern high-frequency interconnects that usually have non-uniform cross-sections caused by discontinuities such as connectors, wire bonds, vias, etc. Hence, the impedance and admittance matrices $\boldsymbol{Z}$ and $\boldsymbol{Y}$ have to be defined so as to take these non-uniformities into account, i.e. $\boldsymbol{Z} = \boldsymbol{Z}(x)$ and $\boldsymbol{Y} = \boldsymbol{Y}(x)$. Therefore, for non-uniform lines, the traditional approach to solving the Telegrapher's Equations, which is valid for uniform lines, cannot be applied since the spatial dependence of $\boldsymbol{Z}$ and $\boldsymbol{Y}$ has to be taken into account when decoupling the equations (3.6) and (3.7). Consequently, a new approach that does not call for the assumption of a longitudinally uniform line is needed.

Instead of differentiating equations (3.6) and (3.7), it has been proposed [C98] to *integrate* them yielding:

$$\boldsymbol{V}(x) = \boldsymbol{V}(l) + \int_{x}^{l} \boldsymbol{Z}(\eta)\boldsymbol{I}(\eta)d\eta \tag{4.1}$$

$$\boldsymbol{I}(x) = \boldsymbol{I}(l) + \int_{x}^{l} \boldsymbol{Y}(\eta)\boldsymbol{V}(\eta)d\eta \tag{4.2}$$

where l is the length of the interconnect. As can be seen, the matrices $\boldsymbol{Z}$ and $\boldsymbol{Y}$ are inside the integration term and dependant on the distance variable; hence no condition is imposed regarding uniformity of the line. In principle, the solution process may continue by substituting (4.2) into (4.1) but this would yield an integral equation for which there is no known solution. In the theory of resonant analysis, the solution to (4.1) and (4.2) is obtained on a discrete basis [C98] as presented in the following section.

4.1.2. Resonant analysis theory

The discretization process starts by dividing the multiconductor line into K sections as shown in Fig. 4.1. The length of the k^{th} section is l_k. It is not necessary to assume that the sections are equal in length or that each section should be longitudinally uniform.

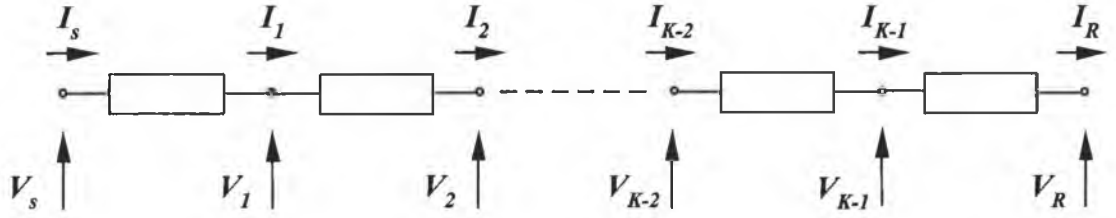


Fig. 4.1. *One-line diagram of a multiconductor line*

In the frequency domain, each section of a model may be represented by an cxact (in the sense that it is a solution of the Telegrapher's equations [C98]) equivalent-π network [C98] as shown in Fig. 4.2.

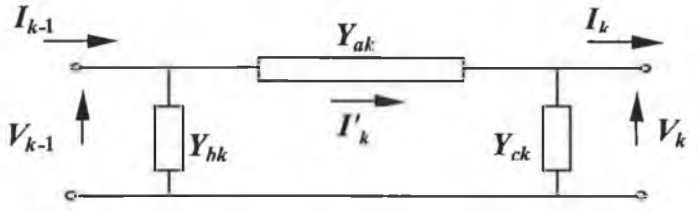


Fig. 4.2. *Multiconductor equivalent-π representation of k^{th} section*

Defining the impedance matrix $\boldsymbol{Z}_{ak}$ by

$$\boldsymbol{Z}_{ak} = [\boldsymbol{Y}_{ak}]^{-1}, \tag{4.3}$$

the system equations for each section may be set up as follows:

$$\begin{aligned} V_k - V_{k-1} &= -Z_{ak} I_k', & k &= 1,2,\ldots,K \\ I_k' - I_{k-1}' &= -(Y_{c,k-1} + Y_{bk}) V_{k-1}, & k &= 2,3,\ldots,K \end{aligned} \tag{4.4}$$

Currents I_k' are defined in Fig. 4.2. Utilising equations (4.4) the system equations may be written in terms of the boundary voltage and current values as:

$$\begin{aligned} V_k &= V_K + \sum_{j=k+1}^{K} Z_{ak} I_j', \quad k = 0,1,\ldots,K-1 \\ I_k' &= I_{K+1}' + \sum_{j=k}^{K} (Y_{b,j+1} + Y_{cj}) V_j, \quad k = 0,1,\ldots,K-1 \end{aligned} \tag{4.5}$$

where $V_0 = V_S$ and $V_K = V_R$ are voltages at the system terminals. The currents at the system terminals ($I_0' = I_S$ and $I_{K+1}' = I_R$) are given by:

$$\begin{aligned} I_S &= I_1' + Y_{b1} V_S \\ I_R &= I_K' - Y_{cK} V_R \end{aligned} \tag{4.6}$$

The transmission line model described by (4.4) or (4.5) is an exact representation of a transmission line, i.e. *no approximation* has been introduced [C98]. The only restriction is that the solution is now available only at the *K+1* nodes of Fig. 4.1, i.e. at discrete points along the line's length.

Equations (4.4) - (4.6) can be written in compact form as the following matrix equation:

$$\begin{bmatrix} I_B \\ V' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_B \\ V' \end{bmatrix} \tag{4.7}$$

where the boundary current vector I_B and the boundary voltage vector V_B in (4.7) are defined as:

$$I_B = \begin{bmatrix} I_S \\ -I_R \end{bmatrix} \quad \text{and} \quad V_B = \begin{bmatrix} V_S \\ V_R \end{bmatrix} \tag{4.8}$$

respectively. The variables I_S, I_R, V_S and V_R are the vectors specified in Fig. 4.1. Voltages at the intermediate multiconductor nodes of Fig. 4.1. are collected in the vector V' defined by:

$$V' = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{K-1} \end{bmatrix} \tag{4.9}$$

The ***ABCD*** matrices can be expressed directly in terms of Z_{ak}, Z_{bk} and Z_{ck} from Fig. 4.2. as defined in Appendix B.

Equation (4.7) may be solved to give an *admittance formulation* that describes the line in terms of its terminal currents and voltages:

$$\boldsymbol{I}_B = \left\{\boldsymbol{A} + \boldsymbol{B}(\boldsymbol{1} - \boldsymbol{D})^{-1}\boldsymbol{C}\right\}\boldsymbol{V}_B \tag{4.10}$$

while the equation

$$\boldsymbol{V}' = (\boldsymbol{1} - \boldsymbol{D})^{-1}\boldsymbol{C}\boldsymbol{V}_B \tag{4.11}$$

provides access to the internal voltages at the *K-1* internal nodes in Fig 4.1. The line response may now be obtained by directly solving equations (4.10) and (4.11). However, with a view to obtaining an efficient transmission line model that can easily be converted to the time domain, equations (4.10) and (4.11) will not be solved directly. Instead, a resonant model based on natural modes of oscillation on the transmission line, as described in the following section, will be employed.

4.1.3. Resonant model

To obtain a resonant model, matrix $\boldsymbol{D}$ in (4.10) and (4.11) is first diagonalised as:

$$\boldsymbol{D} = \boldsymbol{Q}\boldsymbol{\alpha}\boldsymbol{Q}^{-1} \tag{4.12}$$

where matrix $\boldsymbol{\alpha}$ is a diagonal matrix whose elements are the eigenvalues of $\boldsymbol{D}$ and the distribution matrix $\boldsymbol{Q}$ is obtained via the diagonalisation process. After some mathematical manipulation [C98], equations (4.10) and (4.11) are transformed to:

$$\boldsymbol{I}_B = \left\{\boldsymbol{Y}_b + \boldsymbol{Y}_{BB}^{*} + \boldsymbol{P}\boldsymbol{\zeta}\,\boldsymbol{g}\boldsymbol{P}^T\right\}\boldsymbol{V}_B = \boldsymbol{Y}_B\boldsymbol{V}_B \tag{4.13}$$

$$\boldsymbol{V}' = \boldsymbol{Q}\boldsymbol{g}\boldsymbol{P}^T\boldsymbol{V}_B\,. \tag{4.14}$$

Equations (4.13) and (4.14) define the new *resonant transmission line model* for transmission lines. Relative to its boundary terminals, the transmission system is represented by its admittance equation (4.13) while equation (4.14) provides access to voltages at the intermediate points along an interconnect length, if those values are needed for design optimisation.

The important transformation matrix $\boldsymbol{P}$ is computed as:

$$\boldsymbol{P} = [\boldsymbol{p}_1\ \ \boldsymbol{p}_2 \cdots \boldsymbol{p}_n] = [\boldsymbol{Q}^{-1}\boldsymbol{C}]^T \tag{4.15}$$

where $\boldsymbol{p}_i$ is the i^{th} column of $\boldsymbol{P}$. The key property of matrix $\boldsymbol{P}$ is that it is, for most practical cases, independent of frequency.

Diagonal matrices $\boldsymbol{g}$ and $\boldsymbol{\zeta}$ are defined in terms of their diagonal elements as:

$$g_i = (1 - \alpha)^{-1} \tag{4.16}$$

$$\zeta_i = \frac{\boldsymbol{p}_i^T \boldsymbol{x}_i}{\boldsymbol{p}_i^T \boldsymbol{p}_i} \tag{4.17}$$

where $\boldsymbol{x}_i$ is the i^{th} column of matrix $\boldsymbol{X}$ defined as

$$\boldsymbol{X} = [\boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \cdots \;\; \boldsymbol{x}_n] = \boldsymbol{BQ}\,. \tag{4.18}$$

Since $\boldsymbol{\zeta}$ and $\boldsymbol{g}$ are diagonal, the admittance matrix $\boldsymbol{Y}_B$ defined in (4.13) may be decomposed as:

$$\boldsymbol{Y}_B = \boldsymbol{Y}_b + \boldsymbol{Y}''_{BB} + \sum_{i=1}^{n} \zeta_i g_i \boldsymbol{p}_i \boldsymbol{p}_i^T\,. \tag{4.19}$$

The matrices $\boldsymbol{Y}_b$ and $\boldsymbol{Y}''_{BB}$ are defined as

$$\boldsymbol{Y}_b = \boldsymbol{A}_1 \text{ and } \boldsymbol{Y}''_{BB} = \boldsymbol{A}_2 \tag{4.20}$$

where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are the component matrices of $\boldsymbol{A}$ defined in Appendix B.

The block-diagram of this new resonant model may be shown as in Fig. 4.3. For an m-line interconnect network, each component matrix in (4.19) has the dimension $2m \times 2m$ and the number of terms in the summation is $n = m \times (K-1)$. The new model thus represents the interconnect network, relative to its boundary terminals, by a set of $2+m\times(K-1)$ multiterminal admittances, connected in parallel.

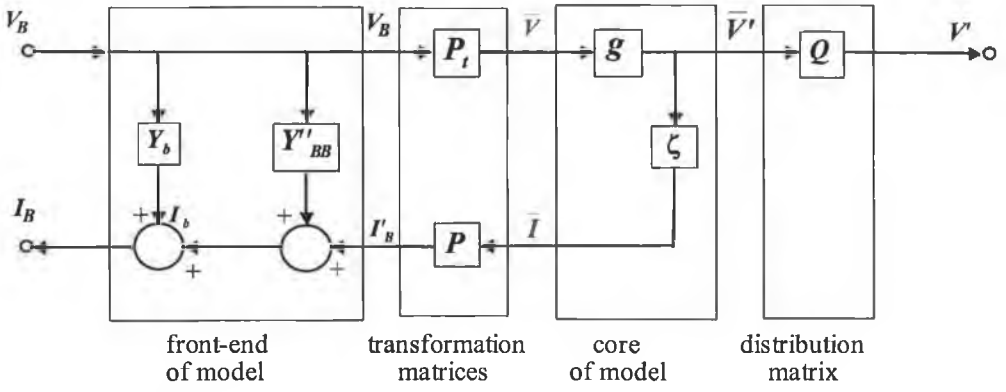


Fig. 4.3. *Resonant model of a transmission line*

The main advantage of this new structure is that it has potential for straightforward conversion into the time domain.

4.1.4. The time-domain conversion

To convert the frequency-domain resonant model described by (4.13) and (4.14) into a time-domain counterpart, the next step is to obtain $\mathcal{Z}$-domain representations for

each of the constituent elements of the model. Usually the transformation and distribution matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ are real and time and frequency independent and there is no need to approximate them. The coefficients of the approximating functions are obtained using auto-regressive moving average (ARMA) modelling [NNA96], [NNA97], where frequency-dependant elements are approximated with $\mathcal{Z}$-domain rational functions as described in the following section.

4.1.4.1. ARMA modelling

In order to perform the conversion to the time domain, the various frequency-dependent elements of the frequency-domain resonant model are approximated with $\mathcal{Z}$-domain transfer functions defined as:

$$APP_{m/n}(z) = \frac{a_0 + a_1 z^{-1} + \ldots + a_m z^{-m}}{1 + b_1 z^{-1} + \ldots + b_n z^{-n}} = \frac{a_0 z^n + a_1 z^{n-1} + \ldots + a_m z^{n-m}}{z^n + b_1 z^{n-1} + \ldots + b_n}, \quad m \le n \tag{4.21}$$

Typically $APP_{m/n}$ is a low-order $\mathcal{Z}$-transfer function and choosing m and n less than 3 will suffice in most cases.

In general, the frequency-dependent element $f(\omega)$ to be approximated is a complex number in which case it may be written as $x + jy$ and equated with an approximating function (4.21)

$$f(\omega) = x + jy = \frac{a_0 z^n + a_1 z^{n-1} + \ldots + a_m z^{n-m}}{z^n + b_1 z^{n-1} + \ldots + b_n} \tag{4.22}$$

Cross multiplying and substituting for z, where z is:

$$z = e^{j\omega\Delta t} = Re + j\,Im \tag{4.23}$$

in (4.22) yields

$$\begin{aligned}(x + jy)\,[(R_e + jI_m)^n + b_1(R_e + jI_m)^{n-1} + \ldots + b_n] = \\ a_0(R_e + jI_m)^n + a_1(R_e + jI_m)^{n-1} + \ldots + a_m(R_e + jI_m)^{n-m}\end{aligned} \tag{4.24}$$

Here Δt is the time step of the model.

Equating the real and imaginary parts on both sides, the following matrix equation is obtained:

$$\boldsymbol{AA}_{2\times(m+n)}\boldsymbol{AB}_{(m+n)\times 1} = \boldsymbol{BB}_{2\times 1} \tag{4.25}$$

where $\boldsymbol{AA}_{2\times(m+n)} = \boldsymbol{AA}(x, y, R_e, I_m)$ and $\boldsymbol{BB}_{2\times 1} = \boldsymbol{BB}(x, y, R_e, I_m)$ and the ARMA coefficients are collected in:

$$\boldsymbol{AB}_{(m+n)\times 1} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \\ b_1 \\ \vdots \\ b_n \end{bmatrix}. \tag{4.26}$$

The matrix equation (4.25) is then solved in a least square sense to obtain the ARMA coefficients $a_0,\ldots,a_m,b_1,\ldots b_n$ of the approximation $APP_{m/n}(z)$ that is a $\mathcal{Z}$-domain representation for each of the constituent frequency-dependant elements of the resonant model. The approximants are then checked for stability. Any poles or zeros that are outside unit circle are reflected back inside thus guaranteeing the stability of all approximations. The significance of obtaining these coefficients is that it is now straightforward to calculate the time-domain response.

4.1.4.2. The choice of approximating functions

The individual elements of the matrices $\boldsymbol{g}$, $\boldsymbol{Y}_b$, $\boldsymbol{\zeta}$ and $\boldsymbol{Y}''_{BB}$ that need to be approximated only require low-order $\mathcal{Z}$-transfer functions. Typically, the maximum order is three. This is shown in [C98] where exact $\mathcal{Z}$-domain expressions are derived for the elements of the matrices $\boldsymbol{g}$, $\boldsymbol{Y}_b$, $\boldsymbol{\zeta}$ and $\boldsymbol{Y}''_{BB}$ for the case of a *lossless* line:

$$g(i,i) = \frac{a_1^g z^{-1}}{1 + b_1^g z^{-1} + b_2^g z^{-2}} \tag{4.27}$$

$$Y_b(i,j) = \begin{cases} a_1^b \dfrac{z^{-1}}{1-z^{-2}}, & i = j \\ -a_1^b \dfrac{z^{-1}}{1-z^{-2}}, & i \neq j \end{cases} \tag{4.28}$$

$$\zeta(i,i) = 2Y''_{BB}(i,i) = a_0^{\zeta}\left\{\frac{1-z^{-1}}{1+z^{-1}}\right\} \tag{4.29}$$

If losses and frequency dependence are to be taken into account, it is recommended [C98] that the order of both the numerator and denominator in (4.27) and (4.28) is increased for approximation of the elements of $\boldsymbol{g}$ and $\boldsymbol{Y}_b$ matrices where the losses are such that this is necessary. For the *lossy* lines encountered in this thesis it is

recommended to model the individual elements of the frequency-dependant matrices $\boldsymbol{g}$, $\boldsymbol{Y}_b$, $\boldsymbol{\zeta}$ and $\boldsymbol{Y}''_{BB}$ as:

$$g(i,i)=\frac{a_1^g z^{-1}+a_2^g z^{-2}}{1+b_1^g z^{-1}+b_2^g z^{-2}+b_3^g z^{-3}} \tag{4.30}$$

$$Y_b(i,j)=\begin{cases}\dfrac{a_1^b z^{-1}+a_2^b z^{-2}}{1+b_1^b z^{-1}+b_2^b z^{-2}+b_3^b z^{-3}}, & i=j\\ -\dfrac{a_1^b z^{-1}+a_2^b z^{-2}}{1+b_1^b z^{-1}+b_2^b z^{-2}+b_3^b z^{-3}}, & i\neq j\end{cases} \tag{4.31}$$

$$\zeta(i,i)=2Y''_{BB}(i,i)=\frac{a_0^\zeta+a_1^\zeta z^{-1}}{1+b_1^\zeta z^{-1}},\quad a_1^\zeta=-a_0^\zeta \tag{4.32}$$

4.1.4.3. Time domain model

After calculating the ARMA coefficients, the line model may be written in the following form:

$$\begin{bmatrix}\boldsymbol{I}_S(z)\\ -\boldsymbol{I}_R(z)\end{bmatrix}=\left[\boldsymbol{Y}_B(z)\right]\begin{bmatrix}\boldsymbol{V}_S(z)\\ \boldsymbol{V}_R(z)\end{bmatrix} \tag{4.33}$$

The elements of the matrix $\boldsymbol{Y}_B(z)$ are calculated from (4.19) after each frequency dependant element is replaced with a suitably chosen approximation of the form (4.21).

Equation (4.33) translates directly to the time domain yielding:

$$\begin{bmatrix}\boldsymbol{i}_S\\ -\boldsymbol{i}_R\end{bmatrix}^{(r)}=\left[\boldsymbol{y}_B\right]^{(r)}\begin{bmatrix}\boldsymbol{v}_S\\ \boldsymbol{v}_R\end{bmatrix}^{(r)}+\begin{bmatrix}\boldsymbol{i}_{his1}\\ \boldsymbol{i}_{his2}\end{bmatrix}^{(r-1)} \tag{4.34}$$

where the superscript 'r' denotes the value at the time t_r. The elements of the $\boldsymbol{y}_B$ matrix are determined from the coefficients of the ARMA models collected in $\boldsymbol{Y}_B(z)$. It is important to note that the history currents $\boldsymbol{i}_{his1}$ and $\boldsymbol{i}_{his2}$ are dependent only on past values of the terminal voltages and currents. Contrary to the convolution approach that requires all past values to calculate the value at the next time point, only a few last values are necessary to obtain $\boldsymbol{i}_{his1}$ and $\boldsymbol{i}_{his2}$. The required number of past points is determined by the chosen order of denominator, n, in the approximation (4.21). As shown in Section 4.1.4.2, this order is usually very low, typically up to three. Thus, calculation of the history currents is not computationally expensive. A detailed derivation for the expressions for $\boldsymbol{i}_{his1}$ and $\boldsymbol{i}_{his2}$ can be found in Appendix C. While the model is derived with a given time-step Δt, this is not a limitation as time-domain interpolation is possible with minimal loss of accuracy. Furthermore, this modelling procedure avoids

the many numerical difficulties and stability issues involved in direct approximation of the *Y* parameters. The whole modelling procedure is illustrated in the following section where the resonant model for a single lossy interconnect with frequency-dependant parameters is obtained.

4.2. Illustrative example – A single lossy frequency-dependant line

In this section the resonant model for the single lossy frequency-dependant interconnect given in Fig 4.4 is derived as an illustrative example. The line of length $l = 0.635\ m$ is described with following p.u.l. parameters: $L =$ 539 nH/m; $G = 0$ S/m; $C =$ 39 pF/m. Skin effect is modelled with a square root dependence as in [O00] i.e.

$$R_{skin}(\omega) \approx R_s(1+j)\sqrt{\omega}\,, \tag{4.35}$$

giving

$$R = R_{dc} + R_s(1+j)\sqrt{\omega} = 0.3691 + 0.0126\,(1+j)\sqrt{\omega}\ \Omega/\text{m}. \tag{4.36}$$

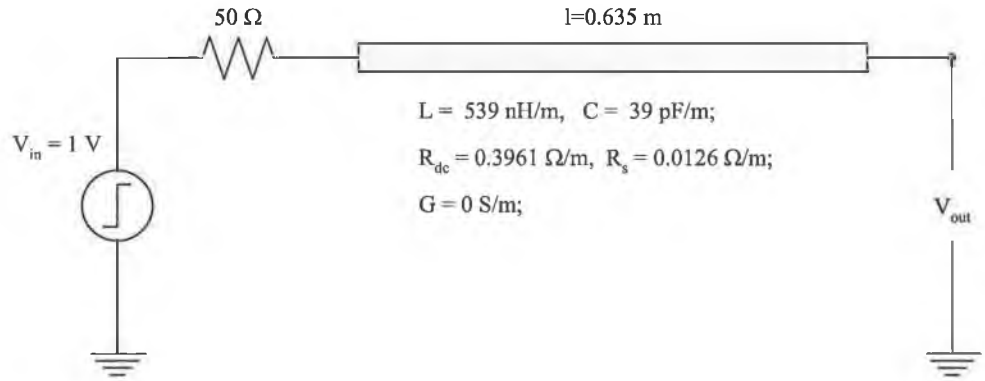


Fig. 4.4. *A single lossy frequency-dependant interconnect line*

The line is divided into $K = 8$ sections and a complete derivation of the resonant model for this interconnect is presented. At the end of the procedure, the response of both the frequency- and time-domain models is compared to confirm the accuracy of the time-domain model.

4.2.1. Deriving a resonant model

In the case of a single uniform line Z_{ak} and Y_{bk} may be written as [C98]:

$$Z_{ak} = Z_0 sinh(l\sqrt{ZY}),\quad k = 1,2,...,K \quad \Rightarrow \quad Z_{ak} = Z_a,\quad \forall k = 1,2,...,K \tag{4.37}$$

$$Y_{bk} = Y_0 tanh(\frac{l\sqrt{ZY}}{2}),\quad k = 1,2,...,K \quad \Rightarrow \quad Y_{bk} = \frac{Y_b}{2}\quad \forall k = 1,2,...,K \tag{4.38}$$

where Z_0 and Y_0 are the characteristic impedance and admittance, respectively and defined as:

$$Z_0 = \sqrt{\frac{Z}{Y}}, \qquad Y_0 = \frac{1}{Z_0} = \sqrt{\frac{Y}{Z}}. \tag{4.39}$$

Therefore, the equivalent-π representation from Fig. 4.2 simplfies to the one in Fig 4.5.

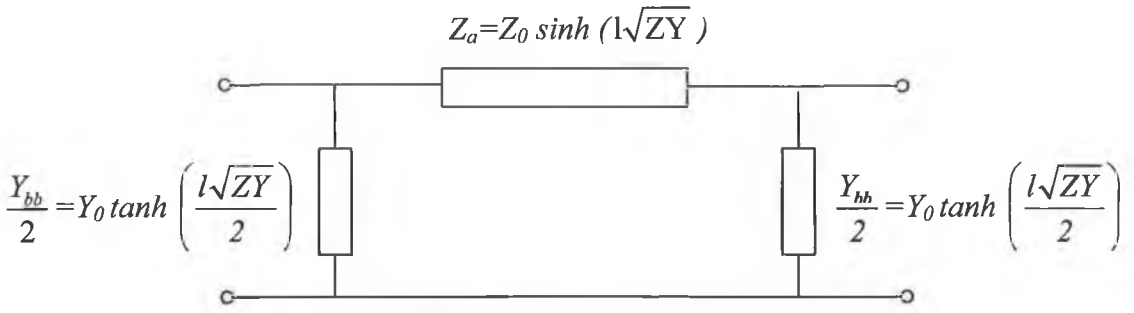


Fig. 4.5. *Equivalent-π representation of each of the K sections of single line*

The expression for $\boldsymbol{Y}_b$ in (4.20) becomes

$$\boldsymbol{Y}_b = \begin{bmatrix} \frac{1}{KZ_a} & -\frac{1}{KZ_a} \\ -\frac{1}{KZ_a} & \frac{1}{KZ_a} \end{bmatrix}, \tag{4.40}$$

and the expression for $\boldsymbol{Y}''_{BB}$ in (4.20) becomes

$$\boldsymbol{Y}''_{BB} = \begin{bmatrix} \frac{Y_{bb}}{2} & 0 \\ 0 & \frac{Y_{bb}}{2} \end{bmatrix}. \tag{4.41}$$

The $\boldsymbol{D}$ matrix defined in Appendix B can be written as:

$$\boldsymbol{D} = -\mu\boldsymbol{R}, \tag{4.42}$$

where μ is a scalar defined as:

$$\mu = Z_a Y_{bb} = 4\sinh^2\left(\frac{l\sqrt{ZY}}{2}\right). \tag{4.43}$$

The elements of the purely real positive-definite matrix $\boldsymbol{R}$ are:

$$\begin{aligned} r_{ij} &= \frac{(K-i)}{K} j \qquad i \geq j \\ r_{ij} &= \frac{(K-j)}{K} i \qquad j > i \end{aligned} \tag{4.44}$$

Since $\boldsymbol{R}$ is symmetrical, it can be diagonalised as:

$$\boldsymbol{R} = \boldsymbol{Q}\boldsymbol{\beta}\boldsymbol{Q}^T, \tag{4.45}$$

where $\boldsymbol{Q}$ is the purely real orthogonal matrix of eigenvectors of $\boldsymbol{R}$ such that $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$. $\boldsymbol{\beta}$ is the diagonal matrix comprising the eigenvalues of $\boldsymbol{R}$. Therefore,

$$\boldsymbol{D} = -\boldsymbol{Q}\mu\boldsymbol{\beta}\boldsymbol{Q}^T. \tag{4.46}$$

Hence, $\boldsymbol{\alpha}$ becomes

$$\boldsymbol{\alpha} = -\mu\boldsymbol{\beta}, \tag{4.47}$$

and thus,

$$\boldsymbol{g} = (\boldsymbol{1} + \mu\boldsymbol{\beta})^{-1} = \begin{bmatrix} g_1 & 0 & \cdots & 0 \\ 0 & g_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{K-1} \end{bmatrix}, \tag{4.48}$$

where

$$g_k = (1 + \mu\beta_k)^{-1} \tag{4.49}$$

and β_k is the k^{th} element of the diagonal matrix β (k^{th} eigenvalue of $\boldsymbol{R}$).

The diagonal matrix $\boldsymbol{\zeta}$ is simply

$$\boldsymbol{\zeta} = \begin{bmatrix} Y_{bb} & 0 & \cdots & 0 \\ 0 & Y_{bb} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Y_{bb} \end{bmatrix} \tag{4.50}$$

In the case of a single uniform line, the definition of $\boldsymbol{C}$ given in Appendix B simplifies to

$$\boldsymbol{C} = \begin{bmatrix} \frac{K-1}{K} & \frac{1}{K} \\ \frac{K-2}{K} & \frac{2}{K} \\ \vdots & \vdots \\ \frac{1}{K} & \frac{K-1}{K} \end{bmatrix} \tag{4.51}$$

By definition, $\boldsymbol{P}^T = \boldsymbol{Q}^{-1}\boldsymbol{C}$ (4.15). Since both $\boldsymbol{Q}$ and $\boldsymbol{C}$ are real, it follows that $\boldsymbol{P}^T$ and $\boldsymbol{P}$ (the transformation matrices of the model in Fig. 4.3) are purely real and independent of frequency for uniform single lines (with or without inclusion of losses).

4.2.2. Conversion to the time domain

Following the procedure described in Section 4.1.4 the next step is to obtain $\mathcal{Z}$-domain representations for the frequency-dependent elements of the frequency-domain model. In the case of a lossy uniform single line, it is necessary to find approximations for the elements of the matrices $\boldsymbol{g}$, $\boldsymbol{\zeta}$, $\boldsymbol{Y}_b$ and $\boldsymbol{Y}''_{BB}$. As suggested in Section 4.1.4.2, the following approximation function has been chosen to approximate $\boldsymbol{g}(i,i)$ and $\boldsymbol{Y}_b(i,i)$:

$$APP_{2/3}(z) = \frac{a_1 z^{-1} + a_2 z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3}} \tag{4.52}$$

The coefficients obtained are shown in Table 4.1 and 4.2.

ARMA coefficients for $\boldsymbol{g}$

Mode i	a_1^i	a_2^i	b_1^i	b_2^i	b_3^i
1	3.0521	0.7035	1.6118	0.9449	0.1835
2	2.7322	0.1852	1.1176	0.7028	0.0871
3	2.2345	-0.3126	0.3959	0.5643	-0.0417
4	1.6301	-0.3982	-0.3283	0.6679	-0.1100
5	1.0130	-0.3225	-1.0317	0.8816	-0.1618
6	0.4828	-0.1924	-1.6510	1.1600	-0.2209
7	0.1258	-0.0650	-2.1325	1.5045	-0.3127

Table 4.1. *ARMA coefficients for g*

ARMA coefficients for $\boldsymbol{Y}_b$

a_1^b	a_2^b	b_1^b	b_2^b	b_3^b
0.0013	-0.0009	-0.6609	-0.5953	0.2897

Table 4.2. *ARMA coefficients for* $\boldsymbol{Y}_b$

The following approximations are used for $\boldsymbol{\zeta}(i,i)$ and $\boldsymbol{Y}''_{BB}(i,i)$

$$APP_{1/1}(z) = \frac{a_0 + a_1 z^{-1}}{1 + b_1 z^{-1}} \tag{4.53}$$

where $a_0 = -a_1$. The coefficients are shown in Table 4.3 and 4.4 Again, the elements for $\boldsymbol{\zeta}(i,i)$ are the same for all modes.

ARMA coefficients for $\boldsymbol{\zeta}(i,i)$, $i = 1,\ldots,7$

Mode	a_0^ζ	a_1^ζ	b_0^ζ
1-7	0.0132	-0.0132	0.7244

Table 4.3. *ARMA coefficients for* $\boldsymbol{\zeta}(i,i)$

ARMA coefficients for Y''_{BB}

a_0^{BB}	a_1^{BB}	b_1^{BB}
0.0066	-0.0066	0.7244

Table 4.4. *ARMA coefficients for* Y''_{BB}

At the conclusion of the linear modelling procedure, the transmission-line model is obtained in the following form:

$$\begin{bmatrix} I_S(z) \\ -I_R(z) \end{bmatrix} = [Y_B(z)] \begin{bmatrix} V_S(z) \\ V_R(z) \end{bmatrix}, \tag{4.54}$$

where

$$Y_B(z) = Y_b(z) + Y''_{BB}(z) + P\zeta gP^T(z). \tag{4.55}$$

Exact expressions for $Y_b(z), Y''_{BB}(z)$ and $P\zeta gP^T(z)$ are given in Appendix C. This format translates directly to the time domain yielding:

$$\begin{bmatrix} i_S \\ -i_R \end{bmatrix}^{(r)} = [y_B]^{(r)} \begin{bmatrix} v_S \\ v_R \end{bmatrix}^{(r)} + \begin{bmatrix} i_{his1} \\ i_{his2} \end{bmatrix}^{(r-1)} \tag{4.56}$$

where the elements of the $[y_B]^{(r)}$ matrix are determined from the coefficients of the ARMA models as derived in Appendix C. The history currents i_{his1} and i_{his2} are dependent only on past values of the terminal voltages and currents and their exact definition may be found in Appendix C.

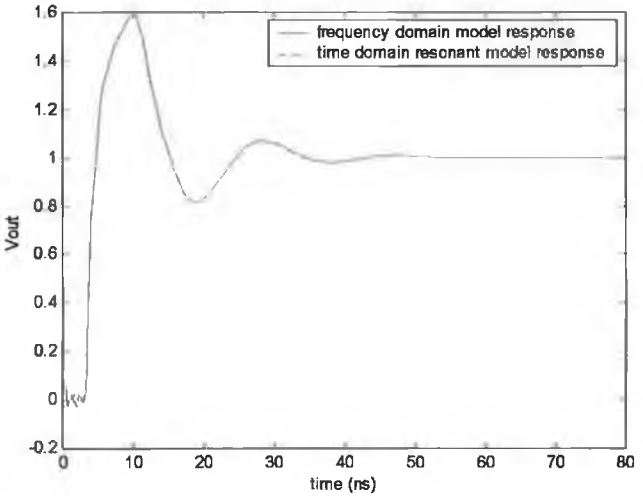


Fig. 4.6. *Output voltage at the open end of the interconnect with step input*

Fig. 4.6 shows a comparison of the boundary output voltage V_{out} calculated from the frequency-domain model and time-domain resonant analysis model at the open end

of the example interconnect. The input voltage is a step function. The response of the new time-domain model arising from resonant analysis is calculated using (4.56) and the following boundary conditions:

i) Step input: $v_S(r) = 1, \quad \forall r \geq 0$

ii) Open circuit at the receiving end: $i_R(r) = 0, \quad \forall r \geq 0$

As evident from this comparison, the responses of frequency-domain model and time-domain resonant analysis model are practically inseparable, thus confirming that the accuracy of the frequency domain model has been preserved in the new time-domain model.

4.3. MOR strategy based on modal elimination

In the previous two sections, a highly accurate resonant model for modelling uniform lossy interconnect with frequency-dependant parameters is presented. In this section, a novel and highly efficient interconnect modelling technique based on exploiting the specific structure of the resonant model is presented. The technique combines the resonant model representation with a model order reduction strategy to produce a highly efficient but nevertheless accurate approach for modelling high-frequency interconnects. The model order reduction strategy based on modal elimination capitalises on the specific structure of the resonant model to enable reduction of an interconnect model.

4.3.1. Introduction

In the resonant model, the relation between the boundary currents and voltages in the frequency domain is given by the admittance equation (4.13) that is repeated here:

$$\boldsymbol{I}_B = \left\{ \boldsymbol{Y}_b + \boldsymbol{Y}''_{BB} + \boldsymbol{P}\boldsymbol{\zeta}\, \boldsymbol{g}\boldsymbol{P}^T \right\} \boldsymbol{V}_B = \boldsymbol{Y}_B \boldsymbol{V}_B . \tag{4.57}$$

Upon closer inspection of this equation, it can be seen that it consists of three parts. The first part, described by the $\boldsymbol{Y}_b$ matrix (4.20), is related to the low-frequency response since $\boldsymbol{Y}_A$ corresponds to the total series impedance (Appendix C). The second part, described by $\boldsymbol{Y}''_{BB}$ (Appendix C) relates to high-frequency response. The third part, $\boldsymbol{P}\boldsymbol{\zeta}\boldsymbol{g}\boldsymbol{P}^T$ corresponds to intermediate frequencies. As was stated in Section 4.2.1, the transformation matrix $\boldsymbol{P}$ and its transpose $\boldsymbol{P}^T$ are purely real and independent of frequency for uniform single lines (with or without inclusion of losses). Hence, only the

product of the matrices $\boldsymbol{\zeta}$ and $\boldsymbol{g}$ in this term is of interest for further analysis. Before proceeding to explain its significance, it is necessary to inspect the nature of $\boldsymbol{\zeta}$ and $\boldsymbol{g}$ in more detail.

4.3.2. Some comments about the nature of ζ

First consider the important theoretical case of a lossless line (R_{dc} and R_s are set to 0) similar to the line in Fig 4.4. For a lossless line, there is an analytical expression for the folding frequency [C98]:

$$f_n = \frac{1}{2l_K\sqrt{LC}} \tag{4.58}$$

If the expression in (4.58) is used to calculate the folding frequency for the example line, the exact value is 1.374 GHz.

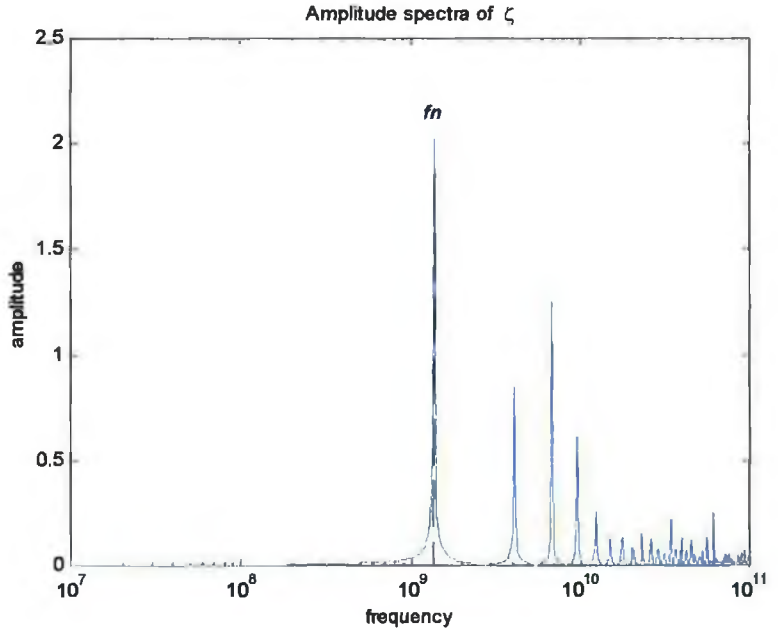


Fig. 4.7. *Amplitude spectra of ζ for a lossless single line*

On the other hand, the amplitude spectra of the elements of the matrix $\boldsymbol{\zeta}$ in the resonant model describing the example lossless line are given in Fig 4.7. The first near-singularity in the amplitude spectra of the elements of the matrix $\boldsymbol{\zeta}$ for this lossless line, occurs at f_n = 1.381 GHz. Obviously, with finite precision computing, the exact frequency cannot be achieved but this result is very close to the exact value of 1.374 GHz. Therefore, the first near-singularity in the amplitude spectra of the elements of the matrix $\boldsymbol{\zeta}$ defines the *folding (or Nyquist) frequency* f_n of the example lossless line.

Hence, it is reasonable to conclude that the folding frequency for the *lossy* line, for which an analytical expression is not available, will be also determined by the first near-singularity in the amplitude spectra of the elements of the matrix $\boldsymbol{\zeta}$.

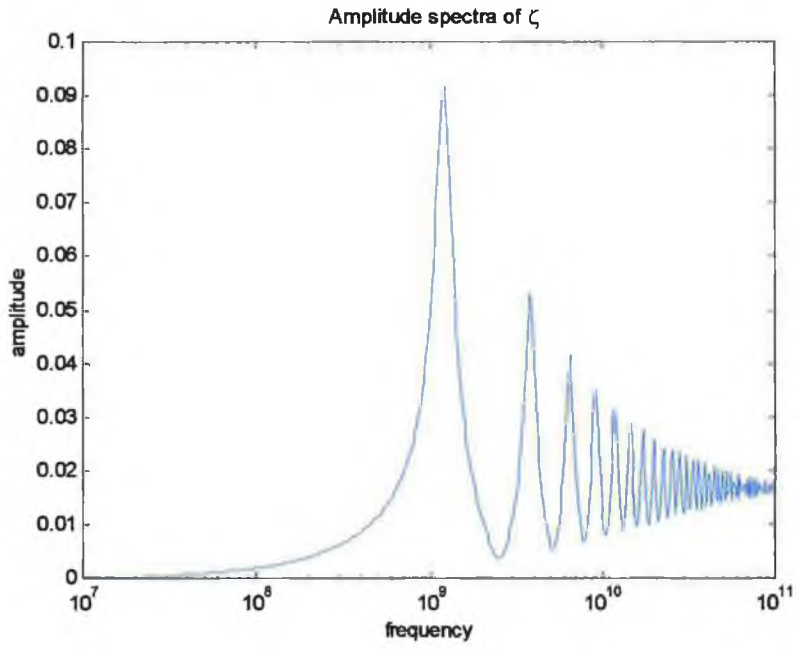


Fig. 4.8. *Amplitude spectra of ζ for a lossy single line*

Fig. 4.8 shows the amplitude spectra of the elements of matrix $\boldsymbol{\zeta}$ for a lossy single line, e.g. R is given by (4.36). It is seen that these elements have a first singularity that defines the folding frequency at f_n=1.087 GHz. As expected, the folding frequency in the case of lossy line is somewhat less than for the previous case of a lossless line since there exist losses on the line and they are taken into account.

4.3.3. The resonant model bandwidth

The folding frequency that is associated with the elements of the matrix $\boldsymbol{\zeta}$ is a very important property of the resonant model. Consider a comparison between the exact amplitude spectra and the spectra obtained from the ARMA approximations for the lossy line, as shown in Fig 4.9. As expected, agreement up to the folding frequency is excellent since the ARMA models are specifically designed to match up to f_n. Similarly, it can be shown that all other frequency-dependent elements (elements of matrices $\boldsymbol{g}, \boldsymbol{Y}_b$ and $\boldsymbol{Y}''_{BB}$) are accurately modelled up to f_n [C98].

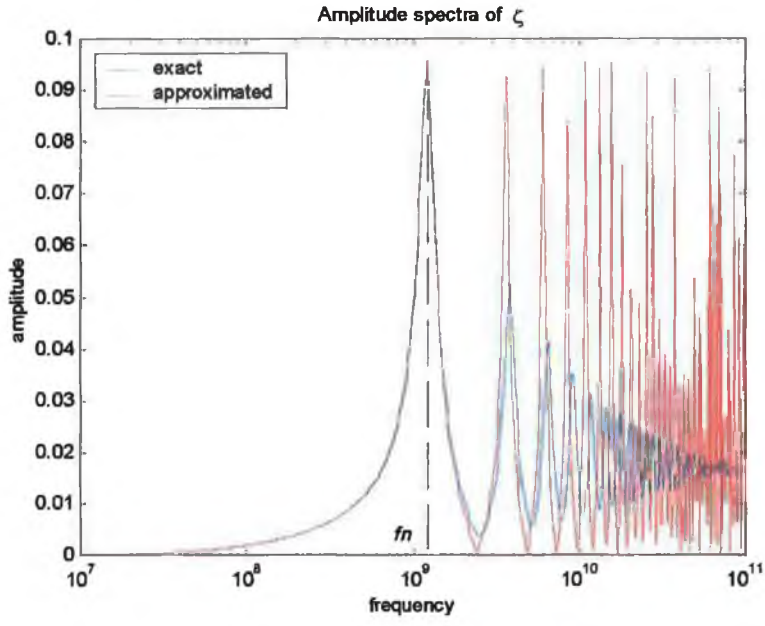


Fig. 4.9. *Comparison between exact and approximated amplitude spectra of* ζ

Therefore, the folding frequency f_n is the upper limit of the resonant model *bandwidth*. If the frequency spectrum of the propagating signal is within the model's bandwidth, the resonant model will accurately model the interconnect behaviour. However, if frequencies that are higher than the folding frequency are present in the system, then the frequency-dependant components may not be properly modelled and hence, errors may arise.

For example, consider the case of the line whose input is a step function that has an infinite frequency spectrum [IJ02], i.e. the maximum frequency present in the system is $f_{max} = \infty$. Consequently, if such signal is to be properly modelled, then the required folding frequency for the interconnect model should be $f_n = f_{max} = \infty$. On the other hand, equation (4.58) implies that the bandwidth of the model is governed by the choice of section length. The shorter section length l_K is chosen, the model's frequency bandwidth becomes wider. Consequently, for the folding frequency to be $f_n = \infty$, the length of the section should be chosen to be infinitely small $(l_K \to 0)$, which is clearly not possible. However, an instantaneous step input that has infinite frequency spectrum is not possible in reality. Instead, any physical signal will have a certain finite albeit short rise time, τ, as illustrated in Fig. 4.10.

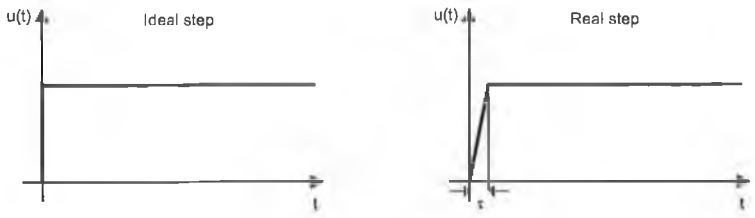


Fig. 4.10. *Ideal and real step input*

From the frequency domain point of view, a finite rise time for a signal means that the frequency spectrum of such a signal will be finite, i.e. f_{max} may be large but still finite. Therefore, when forming the resonant model for an arbitrary interconnect, it will always be possible to choose l_K such that model's folding frequency corresponds to the maximum operational frequency of interest for the designed circuit. The shorter the rise times of the signals that are propagating through the interconnect are, the higher the frequency content of the signal is and thus the smaller l_K will be.

Finally, in agreement with the Sampling Theorem [IJ02], the folding frequency is used to define the time-step of the time-domain model as:

$$\Delta t = \frac{1}{2f_n}, \tag{4.59}$$

In the case of coupled lines when different time steps are involved, linear interpolation is used to combine the transfer functions [C02a]. The lowest folding frequency defines the bandwidth of the resultant time-domain model as all frequency-dependent elements are accurately approximated up to this frequency.

To summarise, the bandwidth of the model may be explicitly estimated since it is determined by the folding frequency of the resonant model. The length chosen for the line sections fixes the folding frequency, which in turn fixes the time step in accordance with the Sampling Theorem. Hence, the choice of section length such that the folding frequency corresponds to the highest frequency of interest ensures an appropriate interconnect representation for the given application.

4.3.4. Some comments about the nature of *g*

Again, consider first the case of a lossless line divided into K sections of equal length l_K. The expression (4.49) for the elements of the matrix $\boldsymbol{g}$ simplifies to [C98]:

$$g_k = \frac{1}{1-4\beta_k \sin^2\left(\frac{\omega l_K \sqrt{LC}}{2}\right)} \tag{4.60}$$

The resonances will occur when the denominator in (4.60) is equal to zero. Therefore, the resonant frequencies are:

$$\omega_k = \frac{2}{l_K\sqrt{LC}} \sin^{-1}\left(\frac{1}{2\sqrt{\beta_k}}\right) \tag{4.61}$$

Setting $k=1$, it can be seen that ω_1 corresponds to the first natural resonant frequency of a short-circuited transmission line ($\Omega_1^{res} = 2\pi / 2l\sqrt{LC}$), ω_2 corresponds to the second natural resonant frequency and so on.

Consider now the lossy single line. The amplitude spectra of the elements of matrix ***g*** are shown in figure 4.11.

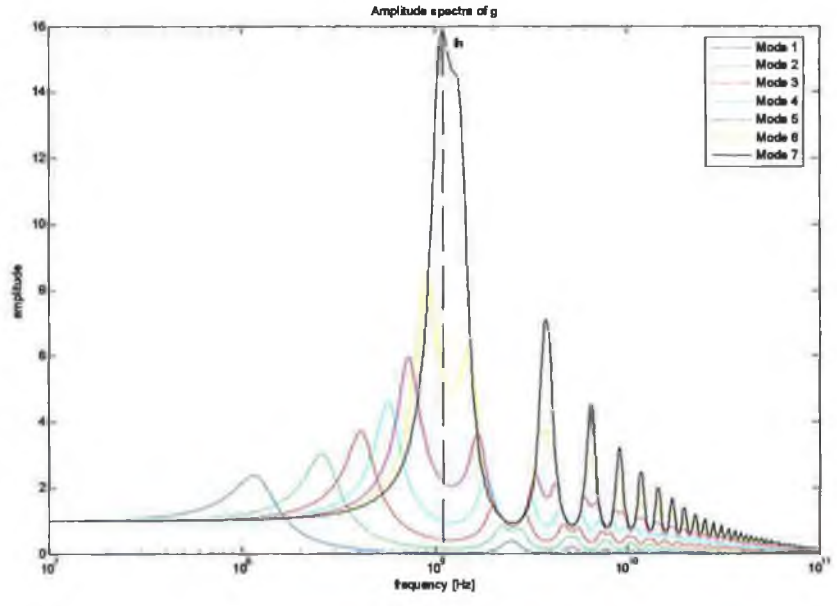


Fig. 4.11. *Amplitude spectra of modal transfer functions for a lossy single line*

The elements of the matrix ***g*** are defined as *modal transfer functions*. Up to the folding frequency, they are seen to have the basic characteristics of lightly-damped low-pass resonant filters [WC97]. The frequencies at which resonances occur define the *natural modes* of oscillation within the model and their numerical values are given in Table 4.5.

Mode	Frequency (GHz)
1	0.116
2	0.262
3	0.423
4	0.580
5	0.738
6	0.904
7	1.087

Table 4.5. *Frequencies of natural oscillation modes for lossy line*

From Fig 4.11. it is seen that up to the Nyquist frequency (1.214 GHz), each mode is characterised by a single resonance after which folding effects occur. The first natural mode has a resonant frequency (0.116 GHz) that corresponds to fundamental resonance. The second resonant frequency (0.262 GHz), corresponds to second-harmonic resonance, the third (0.423 GHz), to third-harmonic resonance, etc. Thus it is clear that the model of Fig. 4.3 is centred around natural modes of oscillation. It should be noted that the natural resonances identified are the short-circuit natural resonances. This is a direct result of the structure of (4.14) which expresses the internal voltages in terms of *both* boundary voltages.

4.3.5. Model order reduction

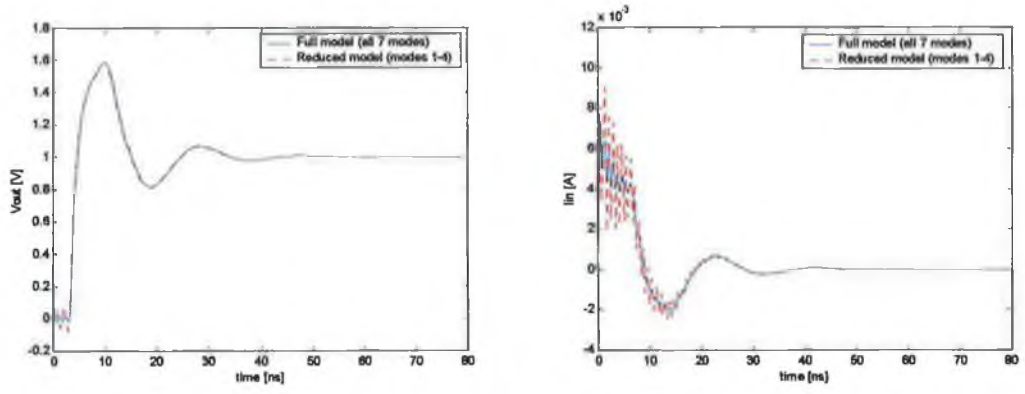
From the discussion presented in Sections 4.3.1 and 4.3.2 it is clear that that matrices $\boldsymbol{\zeta}$, and $\boldsymbol{g}$ represent the *core* of the resonant model. The first near-singularity in the amplitude spectra of the elements of the matrix $\boldsymbol{\zeta}$ defines the folding (Nyquist) frequency f_n. The elements of matrix $\boldsymbol{g}$ identify the natural modes of oscillation of the model. If the highest frequency determining the required bandwidth is smaller than the resonant frequency of a particular mode, the reasonable assumption is that neglecting such a mode will not have a great impact on the accuracy of a model but the size of the model will be reduced thereby yielding a more efficient representation.

The structure of the resonant model is such that it is straightforward to disregard the mode. Neglecting the j^{th} mode is done by simply neglecting the j^{th} term in the summation in equation (4.19). This corresponds to deleting the j^{th} column of $\boldsymbol{P}$ and,

consequently, the j^{th} row of $\boldsymbol{P}^T$, deleting the j^{th} column of $\boldsymbol{Q}$ and deleting the j^{th} rows and columns of the $\boldsymbol{\zeta}$ and $\boldsymbol{g}$ matrices.

4.3.6. Experimental results

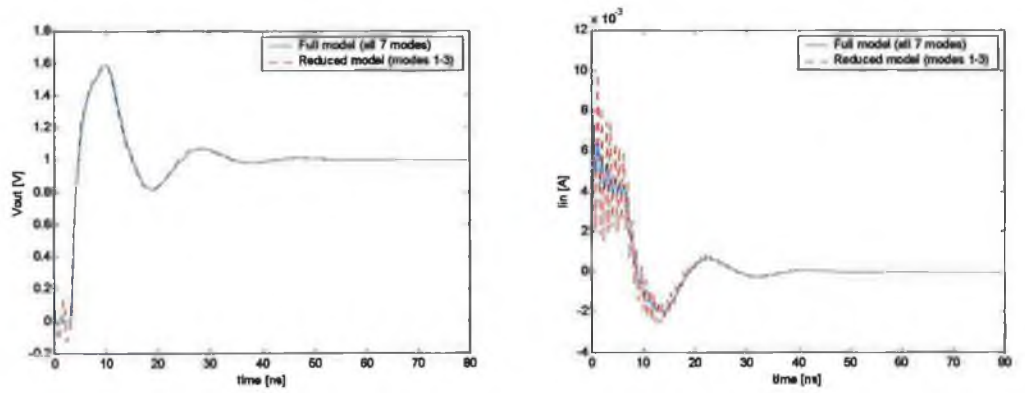
Consider the lossy single transmission line given in Fig. 4.4. Assume that the highest operating frequency will be 0.5 GHz. Upon consulting Table 4.5, in which frequencies of natural oscillation modes are given, it can be seen that only modes 1-4 need to be included in resonant model. Mode 5 is characterised by a resonant frequency of 0.738 GHz and it is reasonable to assume that it and any subsequent modes (6 and 7) will not have much influence.



a) Output voltage V_{out} b) Input current I_{in}

Fig. 4.12. *Reduced model results (4 out of 7 modes)*

Fig 4.12. a) shows the line response (voltage at the open end of a circuit in Fig 4.4) calculated with a model based on the first 4 modes (1-4), compared against the response based on the full model (model base on all 7 modes). As can be seen, the agreement is excellent although the size of the original system has been reduced by 43%.



a) Output voltage V_{out} b) Input current I_{in}

Fig. 4.13. *Reduced model results (3 out of 7 modes)*

Now consider the same situation but with the maximum frequency of interest set to 0.4 GHz. From Table 4.5 it may be concluded that only modes 1-3 need to be included in the model. Fig. 4.13.a) shows the line response calculated with a model based on the first 3 modes (1-3). As may be noted, even with the 53% reduction in the size of the original system, the agreement between the full and the reduced model responses is still very good. For completeness, Appendix E contains diagrams comparing the outputs from reduced models omitting between 1 and 6 modes.

The input currents for both reduction cases (3 out of 7 and 4 out of 7 modes) are shown in the Fig. 4.12.b) and Fig. 4.13.b). What is interesting to note is that there is ripple in both voltage and the current in the reduced model response during the first few nanoseconds. This is due to the finite bandwidth of the model. From Appendix E, it is clear that the inclusion of extra modes improves the quality of the output response.

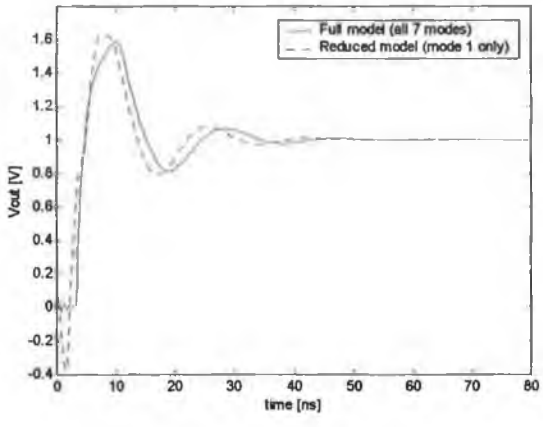


Fig. 4.14. *Reduced model (mode 1 only) response*

Consider now the response of the reduced order model that utilises modes 1-4 as shown in Fig. 4.12.a). What can be noted from the response is that the accuracy of the reduced model is excellent and the only obvious discrepancy between the response of the reduced model and the full model is around initial time. However, this is to be expected. Very high frequency components are introduced if the propagating signal is a step input. Neglecting the "higher frequency" modes (e.g. modes 5 and higher) means that only frequencies up to 0.580 GHz (resonant frequency for mode 4) will be identified by the reduced model. Hence, the discrepancy between the full and the reduced model around the initial time, when the input signal rises from zero to its final value. However, in this particular case, this is not a problem since the design requirement is to capture frequencies up to 0.5 GHz.

4.3.7. Error distribution

All error comparisons presented in this section are made by taking the non-reduced model (all 7 modes taken into account) as the 'exact' value. Hence, the average error is defined as:

$$\varepsilon_{avg} = \frac{1}{N}\sqrt{\sum_{j=0}^{N}(V_{out}^{7}(j) - V_{out}^{i}(j))^2}, \quad i = 1,...,6 \tag{4.62}$$

where $V_{out}^{i}(j)$ is the response calculated at time $t = j\Delta t$, $j = 0,...,N$ by taking the first i modes into account. A bar diagram of the *average error* introduced by neglecting higher modes is shown in Fig 4.15. As can be seen the average error reduces exponentially with the inclusion of extra modes.

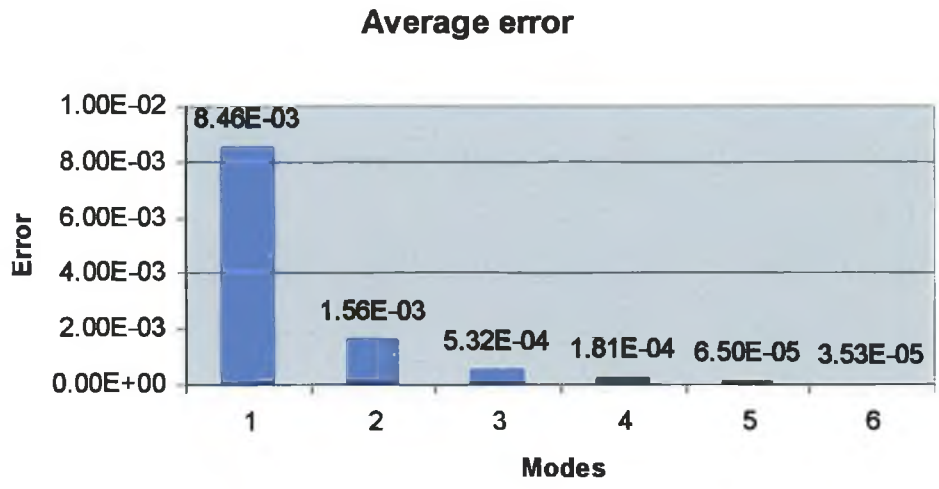


Fig. 4.15. *Average error*

Fig. 4.16 shows the *absolute error* over time where this quantity is defined as:

$$\varepsilon_{abs} = \left|V_{out}^{7} - V_{out}^{i}\right|, \quad i = 1,...,6 \tag{4.63}$$

while Fig. 4.17 shows the relative error

$$\varepsilon_{rel} = \frac{\left(V_{out}^{7} - V_{out}^{i}\right)}{\left|V_{out}^{7}\right|} \times 100\ \%, \quad i = 1,...,6 \tag{4.64}$$

where V_{out}^{i} is the response calculated by taking modes 1 to i into account. Both the absolute and relative errors get smaller as the number of modes taken into account rises.

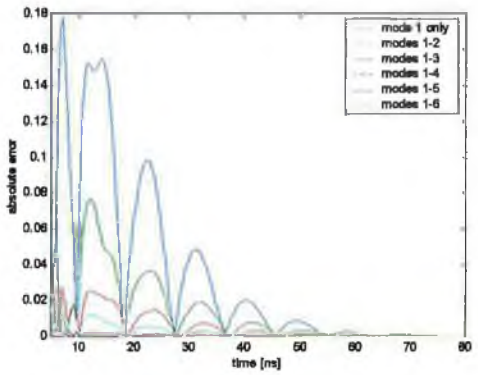


Fig. 4.16. *Absolute error*

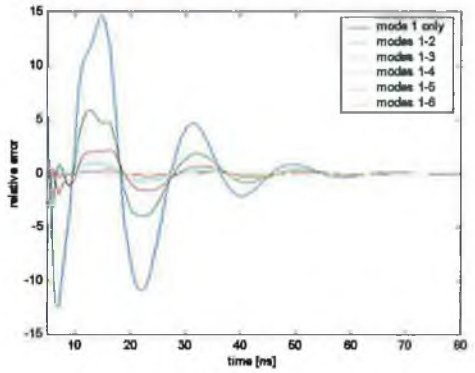


Fig. 4.17. *Relative error*

As can be seen, the accuracy of the reduced model is excellent. Even with a reduced model formed with only one mode, the relative error is less than 15 %.

4.4. Time-domain MOR technique based on the Lanczos process

While the individual ARMA models for each element in the resonant model described in Section 4.1 are of low order, the overall order of the elements of $\boldsymbol{Y_B}(z)$ in (4.54) may be quite high. Consequently, this section suggests a strategy for significantly reducing the order of the model thereby obtaining huge gains in computational efficiency.

4.4.1. Reduced order modelling procedure

The first step involves rearranging the resonant model equations in the $\mathcal{Z}$-domain given in (4.54) into the standard form of a state-space representation. i.e.

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k), \qquad \boldsymbol{A} \in \Re^{n\times n},\ \boldsymbol{x}, \boldsymbol{B} \in \Re^{n},\ \boldsymbol{u} \in \Re \\ y(k) &= \boldsymbol{C}\boldsymbol{x}(k), \qquad\qquad\quad \mathrm{y} \in \Re,\ \boldsymbol{C} \in \Re^{n} \end{aligned} \tag{4.65}$$

The conventional approach is to use the techniques such as the canonical controllability and canonical observability realisations. However, if the matrix $\boldsymbol{A}$ in (4.65) is poorly scaled, this leads to an ill-conditioning problem similar to the one discussed in Section 3.3.6. Therefore, a different approach is needed.

The approach adopted follows from that proposed by Silveira et al. [SEW94] for continuous systems. If the transfer function, $H(z)$, that relates the required system output to the system input may be represented in $\mathcal{Z}$ - domain with the pole-residue representation:

$$H(z) = \sum_{k=1}^{n} \frac{r_k}{z - p_k} \tag{4.66}$$

then the $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ matrices in (4.65) are chosen as:

$$\begin{aligned} \boldsymbol{A} &= diag(p_1 ..... p_n) \\ \boldsymbol{B} &= (\sqrt{|r_1|} ..... \sqrt{|r_n|})^T \\ \boldsymbol{C} &= (sign(r_1)\sqrt{|r_1|} ..... sign(r_n)\sqrt{|r_n|}) \end{aligned} \tag{4.67}$$

For complex conjugate poles, an order 2 state-space representation is formed for each pair of complex conjugate poles and the corresponding 2×2 blocks are inserted into the $\boldsymbol{A}$ matrix. Having formed a well-conditioned state-space realisation, the second step in forming a reduced-order interconnect model is to apply a standard model reduction technique. For the reasons stated in Chapter 3, the Lanczos process [AS00] is deemed suitable technique.

4.4.2. Lanczos process

The Lanczos process [AS00] for model order reduction of a system given in (4.65) may be summarised as follows. Let $\boldsymbol{O}_q$ be the $q \times n$ observability matrix and let $\boldsymbol{R}_q$ be the $n \times q$ reachability matrix defined as:

$$\begin{aligned} \boldsymbol{O}_q &= \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{CA} \\ \vdots \\ \boldsymbol{CA}^{q-1} \end{bmatrix} \\ \boldsymbol{R}_q &= \begin{bmatrix} \boldsymbol{B} & \boldsymbol{AB} & \cdots & \boldsymbol{A}^{q-1}\boldsymbol{B} \end{bmatrix} \end{aligned} \tag{4.68}$$

where q is the order of reduced system. Then an LU factorisation (Appendix A) of the $q \times q$ Hankel matrix $\boldsymbol{H}_q$ defined as:

$$\boldsymbol{H}_q = \boldsymbol{O}_q \boldsymbol{R}_q \tag{4.69}$$

is carried out to obtain matrices $\boldsymbol{L}$ and $\boldsymbol{U}$, i.e.:

$$\boldsymbol{H}_q = \boldsymbol{LU} \tag{4.70}$$

Matrices $\boldsymbol{L}$ and $\boldsymbol{U}$ are used to define projections, $\boldsymbol{\pi}_R$ and $\boldsymbol{\pi}_L$, where

$$\boldsymbol{\pi}_L = \boldsymbol{L}^{-1}\boldsymbol{O}_q, \quad \boldsymbol{\pi}_R = \boldsymbol{R}_q\boldsymbol{U}^{-1}. \tag{4.71}$$

These two projections are then used to define the reduced order matrices:

$$\hat{\boldsymbol{A}} = \boldsymbol{\pi}_L \boldsymbol{A} \boldsymbol{\pi}_R, \quad \hat{\boldsymbol{B}} = \pi_L \boldsymbol{B}, \quad \hat{\boldsymbol{C}} = \boldsymbol{C}\boldsymbol{\pi}_R \tag{4.72}$$

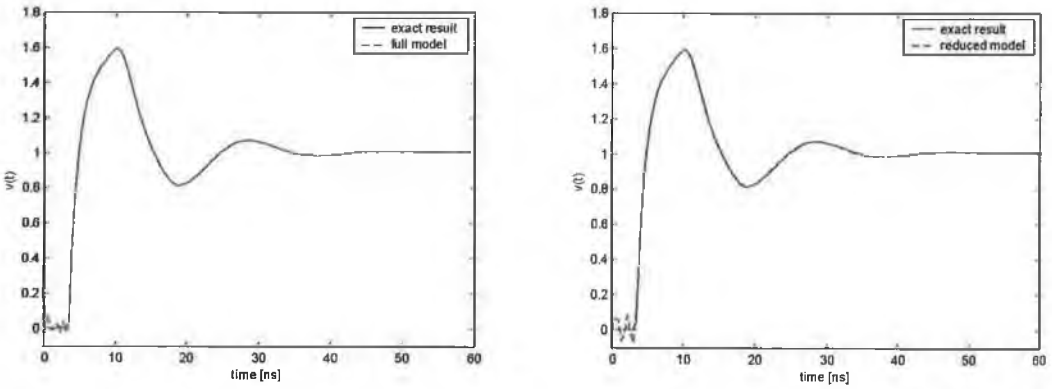
The result is a reduced-order model given by:

$$\begin{aligned} \hat{x}(k+1) &= \hat{A}\hat{x}(k) + \hat{B}u(k), \qquad \hat{A} \in \Re^{q\times q},\ x, B \in \Re^{q},\ u \in \Re \\ y(k) &= \hat{C}\hat{x}(k), \qquad\qquad y \in \Re,\ \hat{C} \in \Re^{q} \end{aligned} \tag{4.73}$$

where the first $2q$ moments of the full model are matched. Since $q \ll n$, the computational cost of solving the reduced system defined with (4.73) is much smaller than directly solving the full order system (4.65).

4.4.3. Illustrative example 1 – A single interconnect

The first example consists of a single interconnect as shown in Fig 4.4. The line is modelled as described in Section 4.2. The order of the input-output transfer function from the full resonant model is 63. The order-reduction process presented in Section 4.4.2 is performed and the order is reduced to 20. Fig. 4.18.a) shows the results from the full resonant model. Fig. 4.18.b) shows the reduced-order model result superimposed on the exact result. As can be seen, the new modelling strategy results in an accurate and efficient model for the interconnect simulation.



a) full model *b) reduced model*

Fig. 4.18. *Open-circuit voltage at receiving-end of the line*

4.4.4. Illustrative example 2 - A coupled interconnect system

The second example is a coupled interconnect system inclusive of skin effect [O00] as shown in Fig. 4.19.

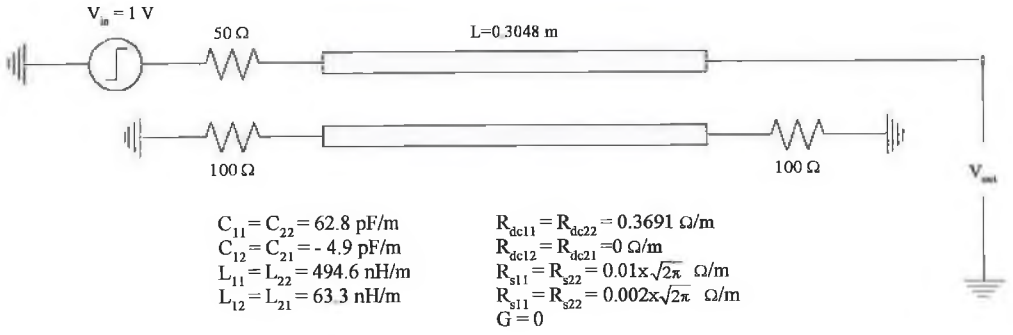
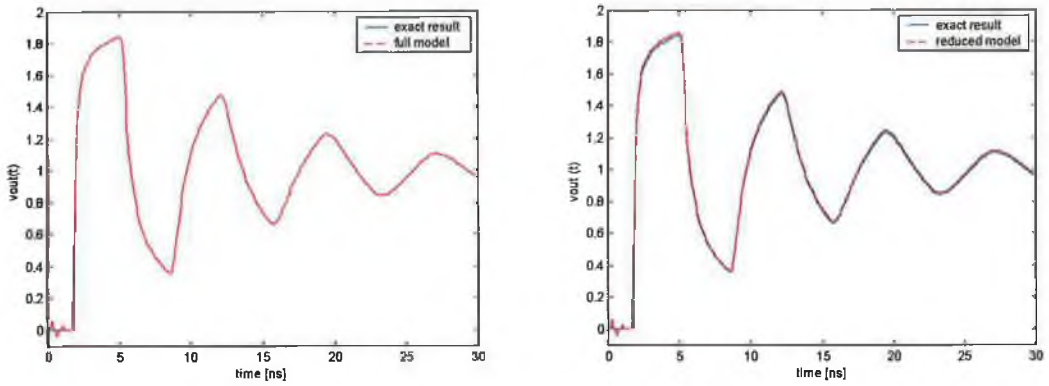


Fig. 4.19. *Coupled transmission line system*

Skin effect is modelled with square root dependence as defined in (4.35). The full resonant model results in a transfer function that is of order 173. The Lanczos process is applied and the order is reduced to 30.



a) full model *b) reduced model*

Fig. 4.20. *Open-circuit voltage*

Figure 4.20.a) compares the result of a *full* time-domain resonant model to the open-circuit voltage result obtained using the frequency-domain model. Figure 4.20.b) compares the result of a *reduced* resonant model to the open-circuit voltage result obtained using a frequency domain model. As can be seen again, excellent accuracy is achieved.

4.5. Conclusion

This chapter has presented two novel modelling techniques for simulation of modern high-frequency interconnects involving resonant analysis and a model order reduction strategy. Initially, the *resonant model* of an interconnect, based on identifying the natural modes of oscillation on the line, is formed. One crucial advantage of the

resonant model is that it does not necessitate the assumption of a longitudinally uniform line. The resonant model prototype, initially formed in the frequency domain, does not introduce any approximations. Hence, it is highly accurate. Since it is formed in the frequency domain it is capable of incorporating the frequency- dependant parameters of a high-speed interconnect line. Hence, the resonant model is capable of handling both uniform and non-uniform interconnects with or without frequency-dependant parameters.

After the highly accurate model prototype is formed in the frequency domain, the model order reduction may be performed with a view to obtaining greater efficiencies. The *first model order reduction strategy* presented in Section 4.3 exploits the modal structure of the resonant model. Depending on the required bandwidth of an interconnect model, the higher modes of a model corresponding to frequencies beyond the required bandwidth may be neglected thus significantly reducing the size of the model but with minimal loss in accuracy. Furthermore, the structure of the resonant model is such that enables straightforward conversion to the time-domain via $\mathcal{Z}$–domain approximation. In addition, the *Lanczos reduction process* in conjunction with a state-space formulation may be applied yielding a significant reduction in the overall model order.

CHAPTER 5

Modelling of Interconnects from a Tabulated Data Set

The contribution presented in this chapter combines in an original manner features from a variety of existing circuit simulation algorithms to result in an efficient interconnect simulation technique for a complex interconnect network described by a tabulated data set. Without loss of generality, the tabulated data set is assumed to be in the form of frequency-dependant *s*-parameters obtained from measurements or rigorous full-wave simulation.

The initial stage in the technique involves a preconditioning of the measured data similar to that proposed in [PB98] for the purposes of ensuring causality of the resultant model for the interconnect network. This is achieved by enforcing the Hilbert Transform relationship that exists between the magnitude function and the phase function of the frequency response for a positive real system. Thereby the causality in the time-domain impulse response corresponding to the measured frequency response is ensured. The impulse response is determined by employing a Reverse Fourier Series approach as proposed in [B95]. In contrast to [B95] where a convolution-based method is used to determine the required transient response, in this contribution, the impulse response is first converted to a $\mathcal{Z}$-domain representation. From this a well-conditioned discrete-time state-space formulation is derived. This enables a judiciously chosen model reduction technique to be employed to reduce the order of the discrete approximation of the system thereby greatly reducing the computational burden involved in obtaining the transient response. The final model achieves both high efficiency and accuracy.

5.1. Introduction

Many interconnect structures for on-chip and chip-to-chip wiring are such that an analytical description of such structures may prove to be a challenging task due to the inhomogeneity of the interconnect geometries involved. In particular, it is difficult to accurately describe interconnects with non-uniform cross-sections caused by

discontinuities such as connectors, vias, wire bonds, redistribution leads, orthogonal lines, insulators with anisotropic dielectric constant, lossy dielectrics, etc [D98]. Very often, an accurate analytical description for these complex interconnect structures is difficult or impossible to obtain. To simulate such interconnects, a designer has to rely on an interconnect description in the form of a tabulated data set. This data is usually in the form of frequency-dependant network parameters such as scattering parameters (s), admittance parameters (y), impedance parameters (z), etc. Section 5.2. gives a brief description of the basic concepts related to these network parameters.

The transient simulation of an interconnect described by a discrete and frequency-dependant data set is not easy task. Schutt-Aine and Mittra [SM88] used a scattering parameter representation in combination with an inverse FFT approach to derive a model for a lossy transmission line that can be linked to non-linear terminations. Apart from the use of time-consuming convolution, the major drawback of this method is the need for an artificial filtering of the s-parameters to reduce the effect of aliasing, as aliasing may result in non-physical behaviour. The non-iterative approach proposed by Dhaene et al. [DMD92], where all coupled ports of the interconnection structure are modelled as extended Thevenin equivalents comprising constant resistances and time-dependant voltage sources, suffers from the same drawback as it uses a bandlimiting window to reduce spurious oscillations in the transient response. A number of authors use rational approximations to the frequency-domain data set in combination with recursive convolution to obtain a time-domain response of an interconnect described with s-parameters. Beyene et al. [BS98] utilise this to form pole-zero models of an arbitrary interconnect, while Neumayer et al. [NSH+02] form a minimal-realisation macromodel. Although the suggested methods do not call for prefiltering of data, both suffer from the ill-conditioning of the large Vandermonde-like matrices involved in obtaining the coefficients of the rational approximations. Furthermore, the number of coefficients in the rational approximation is usually quite high and seriously limits the efficiency of the proposed methods. Silveira et al. [SEW+94] utilise a Truncated Balanced Realisation to address this issue but, as mentioned in Chapter 3, such reduction techniques are unsuitable for the large system models that arise in the technique. Recently, Saraswat et al. [SAN04a], [SAN04b] proposed the reduction of a rational approximation matrix in the frequency domain through a dominant pole-zero approach.

The proposed simulation technique for interconnects described by a tabulated data set abandons the approach of rational approximation of frequency-domain

parameters. Instead, it utilises a Reverse Fourier Series to obtain an approximation in the form of a Finite Impulse Response (FIR) filter in the $\mathcal{Z}$-domain.

5.2. Transmission line description in terms of the network parameters

Consider general two-port network in Fig 5.1. By convention both the input (I_1) and output (I_2) currents flow *into* the 2-port network.

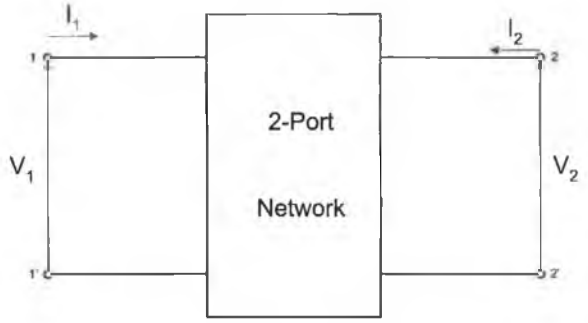


Fig 5.1. *General two-port network*

The parameters that describe the network may be written in the form of *admittance* (*y*-parameters), *impedance* (*z*-parameters), *hybrid* (*h*-parameters), *chain* (*A*- parameters) or *scattering* parameters (*s*-parameters).

5.2.1. The network parameters

The choice of parameter set to be used depends on the specific network at hand. The key factor to consider is the frequency of the signal propagating through the network [HS96].

5.2.1.1. Parameters for low-frequency application

At low frequencies (LF), network analysis may be performed using a LF model represented by either *y*-, *z*- or *h*-parameters that describe the network in terms of a relationship between terminal voltages and currents (I_1, I_2, V_1 and V_2). For example, in terms of *y*-parameters, the two-port network is described by:

$$\begin{aligned} I_1 &= y_{11}V_1 + y_{12}V_2 \\ I_2 &= y_{21}V_1 + y_{22}V_2 \end{aligned} \quad (5.1)$$

To measure these parameters, either a short or open-circuit is required, e.g. to measure the y_{11} parameter , the output (port 2) is short-circuited ($V_2=0$) and after the currents and input voltage are measured, y_{11} and y_{21} may be calculated as:

$$y_{11} = \left.\frac{I_1}{V_1}\right|_{V_2=0} \qquad y_{21} = \left.\frac{I_2}{V_1}\right|_{V_2=0} \tag{5.2}$$

Short circuiting the input ports and repeating the procedure will yield y_{12} and y_{22} as:

$$y_{12} = \left.\frac{I_1}{V_2}\right|_{V_1=0} \qquad y_{22} = \left.\frac{I_2}{V_2}\right|_{V_1=0} \tag{5.3}$$

5.2.1.2. Parameters for high-frequency applications

There are a few practical problems associated with the measurement of *y*-, *z*-, *h*- or *A*-parameters at high frequencies since they require short and open circuits by definition. But at high frequencies (when the wavelength is comparable to the line's dimensions) the lines to/from the measurement system will act as a load to such a system and hence, the condition of a short/open circuit will not be fulfilled. As a rule of thumb, if the circuit operating frequencies are above 100 MHz, a high frequency (HF) model should be used [HS96].

The high-frequency (HF) model utilises the *s*-parameters to model network behaviour. It is based solely on the wave representation where the *power flow* is the property being observed and *not current flow*. More details on s-parameters are given in Section 5.2.2. It should be noted that transformations between all network parameters (*s*-, *y*-, *z*-, *h*- or *A*-parameters) are possible and analytical relationships are readily available, e.g. [HS96], [C92] and [P98]. Therefore, bearing in mind that this thesis is concerned with high-frequency applications, and without loss of generality, from this point forward only *s*-parameters will be considered but the technique developed here may readily be applied to networks characterised by any set of network parameters.

5.2.2. The *s*-parameters

From the theory of transmission lines, it is well known that terminal currents and voltages can be expressed in terms of travelling voltage and current waves [Y90] as:

$$\begin{aligned} V_1 &= V_1^+ + V_1^- \qquad & V_2 &= V_2^+ + V_2^- \\ I_1 &= \frac{V_1^+ - V_1^-}{Z_0} \qquad & I_2 &= \frac{V_2^+ - V_2^-}{Z_0}, \end{aligned} \tag{5.4}$$

where the '+' and '-' superscripts refer to whether the travelling wave is going into or coming out from the two-port network. Z_0 is an arbitrary reference impedance constant.

Relationships (5.4) may now be used to eliminate terminal currents and voltages from (5.1). After some simple mathematical manipulations, one obtains:

$$\begin{aligned} \frac{V_1^-}{\sqrt{Z_0}} &= \frac{f_1(y)}{\sqrt{Z_0}} V_1^+ + \frac{f_2(y)}{\sqrt{Z_0}} V_2^+ \\ \frac{V_2^-}{\sqrt{Z_0}} &= \frac{f_3(y)}{\sqrt{Z_0}} V_1^+ + \frac{f_4(y)}{\sqrt{Z_0}} V_2^+ \end{aligned}, \tag{5.5}$$

where division by $\sqrt{Z_0}$ is preformed for normalisation purposes. Noting that:

$$\begin{aligned} a_1 &= \frac{V_1^+}{\sqrt{Z_0}}, & a_2 &= \frac{V_2^+}{\sqrt{Z_0}} \\ b_1 &= \frac{V_1^-}{\sqrt{Z_0}}, & b_2 &= \frac{V_2^-}{\sqrt{Z_0}} \end{aligned} \tag{5.6}$$

and

$$\begin{aligned} s_{11} &= \frac{f_1(y)}{\sqrt{Z_0}}, & s_{12} &= \frac{f_2(y)}{\sqrt{Z_0}} \\ s_{21} &= \frac{f_3(y)}{\sqrt{Z_0}}, & s_{22} &= \frac{f_4(y)}{\sqrt{Z_0}} \end{aligned} \tag{5.7}$$

equations (5.5) become

$$\begin{aligned} b_1 &= s_{11} a_1 + s_{12} a_2 \\ b_2 &= s_{21} a_1 + s_{22} a_2 \end{aligned}, \tag{5.8}$$

or, in matrix notation:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{5.9}$$

The parameters s_{ij}, $i, j = 1, 2$ are known as *scattering parameters* (*s*-parameters). They are uniquely defined if the impedance level Z_0 is fixed. The important thing to note is that the value of the measured *s*-parameters for the same network will be different if different reference impedance is set. Usually, for interconnect networks, the reference impedance Z_0 is set to 50Ω for reasons explained in Section 3.1.4.

5.2.2.1. The physical interpretation of the *s*-parameters

In case of a 2-port network, s_{11} is the input reflection coefficient and s_{21} is direct gain (attenuation). The parameter s_{22} is the output reflection coefficient and s_{12} is the reverse gain of the network.

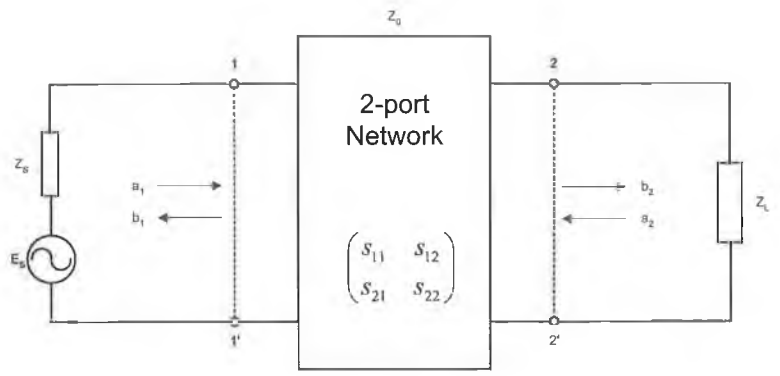


Fig 5.2. *Two-port s-parameters representation*

From (5.6), one may observe that a_1, b_1, a_2 and b_2 are the square roots of the incident and reflected (scattered) powers at ports 1 and 2, respectively. Therefore, the equations (5.8) may be interpreted as the linear relationship between the incident powers (independent variables) and the reflected powers (dependent variables). In that case, the propagation of a signal through a transmission line may be seen as a transfer of the power from the input (port 1-1') to the output (port 2-2') of the 2-port network. Bearing this in mind, the equivalent *s*-parameter representation of a 2-port network may be given as in Fig 5.2.

There is one key difference between the two port *s*-parameter presentation of a network and the representation in Fig. 5.1. The values considered at port 1 are not the current (I_1) and voltage (V_1) but a_1 and b_1, which are the square roots of the powers at the port 1. The situation is similar for port 2-2'. Therefore, the *s*-parameters relate the power at the input to a network to the power at the output and the power flow through the network is the value being observed. This is why the *s*-parameters are suitable for HF network representation.

5.2.2.2. An *n*-port network representation in terms of the *s*-parameters

An *n*-port network may be represented with an *nxn* scattering matrix $\boldsymbol{S}$ defined as:

$$\boldsymbol{S} = \begin{pmatrix} S_{11} & \cdots & S_{1n} \\ \vdots & \ddots & \vdots \\ S_{n1} & \cdots & S_{nn} \end{pmatrix}. \tag{5.10}$$

The elements of the matrix $\boldsymbol{S}$ are the scattering parameters for an *n*-port network and in general, are all frequency-dependent. They may be represented as complex numbers in

terms of either real ($Re\{s_{ij}\}$) and imaginary ($Im\{s_{ij}\}$) parts or in terms of the amplitude (A) and phase ($\angle\varphi$) as in:

$$s_{ij} = Re\{s_{ij}\} + Im\{s_{ij}\} = A\angle\varphi, \quad i,j = 1,\ldots,n \tag{5.11}$$

Now, for a n-port network, equations (5.8) may be written as

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{5.12}$$

or in compact form:

$$\boldsymbol{B} = \boldsymbol{SA} \tag{5.13}$$

where

$$\boldsymbol{B} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \quad \text{and} \quad \boldsymbol{A} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}. \tag{5.14}$$

In equation (5.13), the outgoing waves (matrix $\boldsymbol{B}$) are expressed in terms of the incoming waves (matrix $\boldsymbol{A}$). The wave amplitudes a_n and b_n are related to the currents (I_n) and voltages (V_n) at the port n by the relations

$$a_n = \frac{V_n + Z_0 I_n}{2\sqrt{2Z_0}} \quad \text{and} \quad b_n = \frac{V_n - Z_0 I_n}{2\sqrt{2Z_0}} \tag{5.15}$$

The factor of $\sqrt{2}$ reduces the peak value to an rms or effective value and the factor of $\sqrt{Z_0}$ normalises the amplitude with respect to power. The incoming power (P_{in}) and the outgoing power (P_{out}) at the port n are defined as:

$$P_{in}^n = \boldsymbol{a}_n \boldsymbol{a}_n^* \quad \text{and} \quad P_{out}^n = \boldsymbol{b}_n \boldsymbol{b}_n^* \tag{5.16}$$

Therefore, the s-parameters may be interpreted as fixed electrical properties of an n-port network that describe how energy couples between each pair of ports of the circuit.

5.2.2.3. Measurement of the *s*-parameters

For the measurement of y-, z-, h- or A-parameters, short and open circuits are required by definition. However, at high frequencies, short and open circuit currents and voltages are very difficult to measure exactly. In addition, most active devices and circuits are not open- or short-circuit stable. Therefore, in high frequency circuit analysis, it is desirable to obtain the system description in terms of parameters that do not require short and open circuits for their measurement.

Consider, now, a standard two-port network as in Fig 5.2 described in terms of *s*-parameters. It is connected to a generator with a source impedance Z_S and to a load Z_L. If the network is connected to a load impedance Z_L equal to reference impedance Z_0 $(Z_L = Z_0)$, then there is no power reflected into the network, i.e. $a_2=0$, and hence the parameters s_{11} and s_{21} may be obtained as:

$$s_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0} \qquad s_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0} \tag{5.17}$$

Interchanging the positions of port 1-1' and 2-2' in the measurement set up, $a_1=0$ and s_{12} and s_{22} may be obtained.

The important thing to note here is that the measurement of *s*-parameters does not require open or short-circuit terminal ports. Hence, the *s*-parameter description of an interconnect network may be obtained with reasonable accuracy at high-frequencies.

5.3. Formation of a discrete-time representation from a data set

The description of a high-frequency interconnect network in terms of *s*-parameters is very useful since *s*-parameters depend only on the networks' electrical characteristics and are not influenced by voltages at terminations. Secondly, as previously stated, their *accurate* measurement at very high frequencies is possible. Thirdly, since any *s*- parameter is the ratio of reflected/incident power, the magnitude of a *s*-parameter is always *less than 1*, i.e. scattering parameters remain bounded and stable. On the other hand, admittance (*y*) or impedance (*z*) parameters can become singular at the resonant frequencies of the network in question. Therefore, the *s*-parameters are chosen as a preferred description of an arbitrary complex interconnect network at high-frequencies.

5.3.1. Enforcement of causality conditions

The values of *s*-parameters are frequency-dependant values due to skin effect, proximity effect and edge effects. Hence, from this point forward, the *s*-parameter data set will be assumed to be in the form of a set of frequency-domain values where $H(\omega)$ denotes the value at the frequency ω.

For the case of data provided by measurement, it is necessary to ensure that errors due to noise or systematic errors do not lead to a non-causal impulse response. Non-causality indicates non-physical behaviour and is inappropriate for interconnect

models. Consequently, for a measured frequency response, it is necessary to precondition the data. To this end, Perry and Brazil [PB98] proposed the Hilbert Transform relationship:

$$\phi(\omega) = \frac{-1}{\pi} \int_{-\infty}^{\infty} \frac{\alpha(\xi)}{\omega - \xi} d\xi \tag{5.18}$$

This relationship relates the phase response of a *positive real* filter to its magnitude response. By enforcing this relationship, causality of the impulse response is ensured. However, because the frequency response is only known over a narrow range of frequencies (between ω_l and ω_h), a reduction in the limits of integration is required.

$$\phi(\omega) = \frac{-1}{\pi} \int_{\omega_l}^{\omega_h} \frac{\alpha(\xi)}{\omega - \xi} d\xi \tag{5.19}$$

The integral may be interpreted as a convolution:

$$\phi(\omega) = \alpha(\omega) * \frac{-1}{\pi\omega} \tag{5.20}$$

Equation (5.20) may be implemented numerically in an efficient manner using the Fast Fourier Transform as described in [PB97]:

$$\phi(\omega) = IFFT\{ FFT(\alpha(\omega))(-jsign(v))\}. \tag{5.21}$$

$\phi(\omega)$ is the phase of the tabulated data set. $|H(\omega)|$ is the magnitude response of the measured frequency domain data and $\alpha(\omega) = \ln|H(\omega)|$. v is the new transform-domain variable and $\{-jsign(v)\}$ is the analytical Fourier Transform of the $-1/\pi\omega$ term.

As stated above, the Hilbert Transform applies to positive real systems. However, scattering parameters are bounded between –1 and +1 and reflection scattering parameters are rarely positive real numbers. Hence, the relationship in (5.18) may not be directly applied. To overcome this, the remedy presented in [PB98] is employed whereby an offset of one is applied to the scattering parameters. The resultant offset parameters are thus positive real functions. The phase of the *s*-parameters is then determined from (5.21) and the offset is removed. In this manner, it is possible to bound the parameters to ensure that a causal impulse response is obtained and that passivity is maintained or enforced.

5.3.2. Determination of the impulse response

Having ensured that the initial set of frequency–domain data describes a physically realisable (causal) system, the next stage involves determining the

corresponding impulse response. To this end, the following discrete-time Fourier Transform pair [B95] is proposed for use:

$$h(nT)=\frac{1}{2\pi}\int_{-\omega_m}^{\omega_m} H(\omega)e^{jn\omega T}d\omega \tag{5.22}$$

$$H(\omega)=T\sum_{n=0}^{\infty} h(nT)e^{-jn\omega T} \tag{5.23}$$

where ω_m is maximum frequency of interest and $T = \pi/\omega_m$.

Two points are worth noting in relation to (5.22) and (5.23). Firstly, note the change in the scaling factors when compared to the traditional Fourier Series. The change in scaling factors is introduced to enable *h(nT)* to limit to the continuous impulse response as ω_m tends to infinity. Secondly, an exponent sign-change is introduced. This sign-change is necessary to maintain causality of the time-domain samples (the opposite sign in the exponent would lead to anti-causal behaviour in the time domain, i.e. samples in the time domain would be zero-valued for positive time).

Let the measured response consist of (N+1) equally-spaced samples of $H(\omega)$ in the frequency range [0, ω_m]. The first sample corresponds to $\omega_0 = 0$ and the last sample corresponds to $\omega_N = \omega_m$. In order to ensure a real-valued time-domain response, the condition of Hermitean symmetry is assumed, i.e.

$$H(\omega)=H^*(-\omega). \tag{5.24}$$

Thus the formula in (5.22) may be written as:

$$h(nT)=\frac{1}{2\pi}\int_{0}^{\omega_m}\left[H^*(\omega)e^{-jn\omega T}+H(\omega)e^{jn\omega T}\right]d\omega=\frac{1}{2\pi}\int_{0}^{\omega_m}F(\omega)d\omega, \tag{5.25}$$

where

$$F(\omega)=H^*(\omega)e^{-jn\omega T}+H(\omega)e^{jn\omega T}. \tag{5.26}$$

The integral in (5.25) on the interval [0, ω_m] may be written as a sum of integrals on intervals [ω_{k-1}, ω_k] as:

$$\begin{aligned} h(nT)&=\frac{1}{2\pi}\int_{0}^{\omega_m}F(\omega)d\omega=\frac{1}{2\pi}\left[\int_{0}^{\omega_1}F(\omega)d\omega+\int_{\omega_1}^{\omega_2}F(\omega)d\omega+\ldots+\int_{\omega_{N-1}}^{\omega_N}F(\omega)d\omega\right] \\ &=\frac{1}{2\pi}\sum_{k=1}^{N}\int_{\omega_{k-1}}^{\omega_k}F(\omega)d\omega \end{aligned} \tag{5.27}$$

To numerically calculate the integrals in (5.27), the trapezoidal rule of integration given as:

$$\int_{x_1}^{x_2} f(x)dx = \frac{\Delta x}{2}[f(x_1)+f(x_2)]-O((\Delta x)^3) \tag{5.28}$$

may be applied yielding

$$\begin{aligned} h(nT) &= \frac{1}{2\pi}\sum_{k=1}^{N}\int_{\omega_{k-1}}^{\omega_k} F(\omega)d\omega \approx \frac{1}{2\pi}\sum_{k=1}^{N}\frac{\Delta\omega}{2}[F(\omega_{k-1})+F(\omega_{k-1})] \\ &= \frac{\Delta\omega}{4\pi}\sum_{k=1}^{N}[F(\omega_{k-1})+F(\omega_{k-1})]. \end{aligned} \tag{5.29}$$

Thus inserting (5.26) into equation (5.29) gives:

$$h(nT) = \frac{\Delta\omega}{4\pi}\sum_{k=1}^{N}\left[H^*(\omega_{k-1})e^{-jn\omega_{k-1}T} + H(\omega_{k-1})e^{jn\omega_{k-1}T} + H^*(\omega_k)e^{-jn\omega_k T} + H(\omega_k)e^{jn\omega_k T}\right] \tag{5.30}$$

This enables the calculation of $2N$ samples of the impulse response $h(nT)$. The developed formula in (5.30) relates a continuous periodic function of frequency to a discrete real-valued function in the time domain up to some specified boundary frequency ω_m. This frequency is the highest frequency at which the s-parameters were measured/simulated. It is very important to choose the frequency ω_m such that the spectral energy beyond ω_m is relatively small. If this is not the case, the errors will arise in the simulated transient response.

5.3.3. Formation of the $\mathcal{Z}$-domain representation

The determination of an FIR filter corresponding to the impulse response is a trivial task as it is well-known that the coefficients of an FIR filter correspond to its impulse response, i.e.

$$H(z) = \sum_{k=0}^{2N-1} h(kT)z^{-k} \tag{5.31}$$

Hence, an FIR filter representation for each element of the descriptor matrix may be directly determined.

5.4. Model reduction procedure

Having obtained the impulse response (5.30) and consequently, an FIR filter representation (5.31), it is possible to use it directly for the purposes of transient analysis. This can be done via inverse $\mathcal{Z}$-transform techniques [OSB99], [IJ02] or by

employing the causal convolution approach as advocated in [B95]. However, in this contribution, a model reduction technique is applied to greatly improve the efficiency of the resultant interconnect system model.

5.4.1. Formation of a well-conditioned state-space representation

To enable a reduction process to be applied it is necessary to convert the required $\mathcal{Z}$-domain representation for the system to a standard state-space format:

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{F}\boldsymbol{x}(k) + \boldsymbol{G}u(k) \qquad & \boldsymbol{F} \in \Re^{n\times n}, \boldsymbol{x}, \boldsymbol{G} \in \Re^{n},\ u \in \Re \\ y(k) &= \boldsymbol{H}\boldsymbol{x}(k) + \boldsymbol{D}u(k) \qquad & y \in \Re,\ \boldsymbol{H} \in \Re^{1\times n} \end{aligned} \tag{5.32}$$

As in Section 4.4.1, suppose that the transfer function, $TF(z)$, that relates the required network output to the network input may be represented as:

$$TF(z) = r_{\infty} + \sum_{k=1}^{n} \frac{r_k}{z - p_k} \tag{5.33}$$

where $TF(z)$ is the required transfer function formed from the individual descriptor parameters, P is the number of poles and r_k is the residue corresponding to the k^{th} pole, p_k. Then the $\boldsymbol{F}$ matrix in (5.32) is chosen as:

$$\boldsymbol{F} = diag(p_1 \ldots\ldots p_n) \tag{5.34}$$

and the $\boldsymbol{G}$ and $\boldsymbol{H}$ matrices are chosen as:

$$\begin{aligned} \boldsymbol{G} &= (\sqrt{|r_1|} \ldots\ldots \sqrt{|r_n|})^{\mathrm{T}} \\ \boldsymbol{H} &= (sign(r_1)\sqrt{|r_1|} \cdots sign(r_n)\sqrt{|r_n|}) \end{aligned} \tag{5.35}$$

The $\boldsymbol{D}$ matrix equals r_{∞}. Again for complex conjugate poles, an order 2 state-space representation is formed for each pair of complex conjugate poles and the corresponding 2×2 blocks are inserted into the $\boldsymbol{F}$ matrix. To ensure the stability of the method, any poles that are outside the unit circle are eliminated.

5.4.2. Laguerre model reduction

Having formed a well-conditioned and stable state-space realisation, the next step in forming a reduced-order network model is to apply a suitable model reduction technique. The particular procedure chosen here is the model reduction technique based on the Laguerre polynomial expansion as introduced in [CBK+02].

5.4.2.1. Laguerre polynomials

The *Laguerre polynomials* L_n $([0,\infty) \rightarrow \Re)$ are polynomials defined as:

$$L_n(t) = \frac{e^t}{n!}\frac{d^n}{dt^n}(e^{-t}t^n), \qquad i = 0,1,2,... \tag{5.36}$$

These polynomials form a complete orthogonal set on the interval $t \in [0,\infty)$ with respect to the weighting function , e^{-t}, i.e.

$$\int_0^\infty e^{-t} L_m(t) L_n(t) dt = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \tag{5.37}$$

The key property of the k^{th} order Laguerre polynomial is that it serves as the optimal k^{th} order approximant (5.38) to the impulse response $x(t)$ of the given system.

$$\hat{x}_k(t) = \hat{c}_0 L_0(t) + \hat{c}_1 L_1(t) + ... + \hat{c}_k L_k(t), \qquad c_i \in \Re,\ i = 0,1,.. \tag{5.38}$$

The optimality is defined in the sense of minimising an exponentially weighted error *ERR:*

$$ERR = \int_0^\infty e^{-t}[x(t) - x_k(t)]^2 dt \tag{5.39}$$

This results in errors close in time to the point of application of the signal being weighted heavily. This, of course, is appropriate to most high-speed applications when signal transitions occur shortly after impulse excitation [CBK+02]. Hence, the employment of a Laguerre model reduction scheme is deemed appropriate in the current context and, for completeness, is reviewed briefly in the next section.

5.4.2.2. Laguerre model reduction scheme

Consider a system:

$$\begin{aligned} \frac{d\boldsymbol{x}(t)}{dt} &= \boldsymbol{F}\boldsymbol{x}(t) + \boldsymbol{G}u(t) \qquad \boldsymbol{F} \in \Re^{n\times n}, \boldsymbol{x}, \boldsymbol{G} \in \Re^n,\ u \in \Re \\ y(t) &= \boldsymbol{H}\boldsymbol{x}(t) + Du(t) \qquad y \in \Re,\ \boldsymbol{H} \in \Re^{1\times n} \end{aligned} . \tag{5.40}$$

Explicit moment matching MOR techniques (AWE and CFH) and the Krylov subspace techniques (Arnoldi and Lanczos) as described in Chapter 3 concentrate on approximating the *frequency-domain* transfer function of the original time-domain system described by (5.40).

In contrast, the Laguerre model reduction technique approximates the *time-domain* system impulse response given by:

$$x(t) = e^{-Ft}\boldsymbol{G} \tag{5.41}$$

with the k^{th} order Laguerre approximation $\hat{x}_k(t)$

$$\hat{x}_k(t) = \sum_{i=0}^{k} \hat{c}_i L_i(t). \tag{5.42}$$

Let

$$\tilde{\boldsymbol{A}} = (\boldsymbol{I} + \boldsymbol{F})^{-1}\boldsymbol{F} \quad \text{and} \quad \tilde{\boldsymbol{B}} = (\boldsymbol{I} + \boldsymbol{F})^{-1}\boldsymbol{G} \tag{5.43}$$

then, the coefficients $\hat{c}_i$ in (5.42) may be obtained as:

$$\begin{aligned} \hat{c}_0 &= (\boldsymbol{I} + \tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{B}} \\ \hat{c}_i &= \frac{1}{i}(\boldsymbol{I} + \tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{A}}\hat{c}_{i-1} \end{aligned} \tag{5.44}$$

Now, for the model order reduction purpose one may define the matrix $\boldsymbol{P}_K$ as:

$$\boldsymbol{P}_K = [\tilde{\boldsymbol{B}} \;\; \tilde{\boldsymbol{A}}\tilde{\boldsymbol{B}} \;\; \cdots \tilde{\boldsymbol{A}}^K\tilde{\boldsymbol{B}}] \tag{5.45}$$

Then, the expression for the K^{th}-order approximation of the impulse response, $x_K(t)$, may be noted as:

$$x_K(t) = \boldsymbol{P}_K \begin{bmatrix} \frac{1}{0!}L_0(t) \\ \frac{1}{1!}L_1(t) \\ \vdots \\ \frac{1}{K!}L_K(t) \end{bmatrix} \tag{5.46}$$

Thus $x_K(t)$ lies in the span of the columns of the matrix $\boldsymbol{P}_K$ for all t.

In light of this, the model reduction scheme projects the full state-space of the system onto the span of the columns of $\boldsymbol{P}_K$. A QR factorisation of $\boldsymbol{P}_K$ is first performed resulting in:

$$\boldsymbol{P}_K = \boldsymbol{Q}_K \boldsymbol{R}_K. \tag{5.47}$$

Subsequently, the reduced-order model of the original system described by (5.32) is given by:

$$\begin{aligned} \hat{\boldsymbol{x}}(k+1) &= \hat{\boldsymbol{F}}\hat{\boldsymbol{x}}(k) + \hat{\boldsymbol{G}}\boldsymbol{u}(k) \\ \overline{y}(k) &= \hat{\boldsymbol{H}}\hat{\boldsymbol{x}}(k) + \boldsymbol{D}\boldsymbol{u}(k) \end{aligned}, \tag{5.48}$$

where

$$\boldsymbol{x}(k) = \boldsymbol{Q}_K\hat{\boldsymbol{x}}(k), \; \hat{\boldsymbol{F}} = \boldsymbol{Q}_K^T\boldsymbol{F}\boldsymbol{Q}_K, \; \hat{\boldsymbol{G}} = \boldsymbol{Q}_K^T\boldsymbol{G} \text{ and } \hat{\boldsymbol{H}} = \boldsymbol{H}\boldsymbol{Q}_K. \tag{5.49}$$

$\boldsymbol{Q}_K^T$ is the transpose of the matrix $\boldsymbol{Q}_K$. The reduced order model (5.48) is passive if the original system (5.32) is passive owing to the orthogonality of $\boldsymbol{Q}_K$ [CBK+02], [OCP97].

Having formed a well-conditioned passive reduced-order model for the electrical network characterised by measured or simulated data, it is now possible to perform the numerical calculations in an efficient manner in order to obtain the transient response of given network ($\bar{y}(k) \approx y(k)$).

5.5. Experimental results

The proposed novel methodology for simulating interconnect networks from measured or simulated data has been tested on the two network topologies given in Fig 5.3 and Fig 5.8 respectively. The findings will confirm the efficacy of the proposed time-domain model.

5.5.1. Illustrative example 1 – The simulated data

The first example is the idealised low-pass filter structure that was also employed in [B95]. Initially, the transmission lines are assumed to be ideal with the characteristic impedance values given in Fig. 5.3. First, the structure was terminated with an impedance of 50 Ω (Fig. 5.3) in order to obtain the scattering parameters that describe this low-pass filter network.

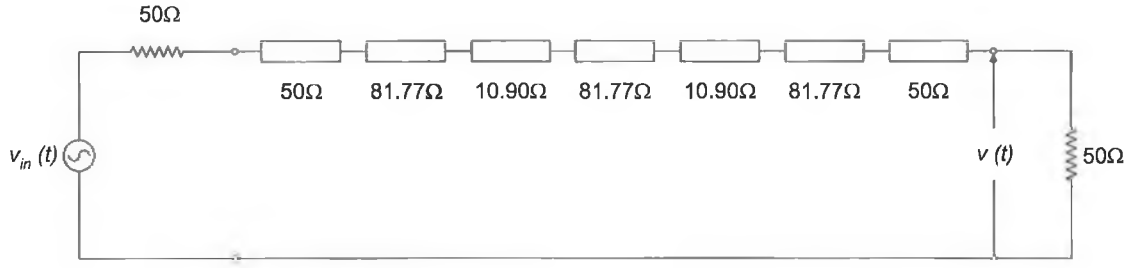


Fig. 5.3. *Sample lossless low-pass filter network setup for obtaining scattering parameters (All lines are 90° at 2.5GHz)*

After obtaining the values of the *s*-parameters, the 50Ω termination is removed (Fig 5.4) and the open-circuit response of this lossless network when the input is a unit step is calculated.

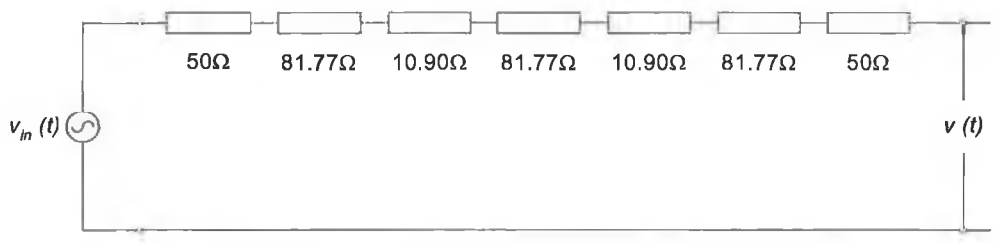


Fig. 5.4. *Sample lossless low-pass filter network with open end for transient analysis*

Fig 5.5 shows the time-domain step response (dashed line) superimposed on the exact frequency-domain model response (dashed line).

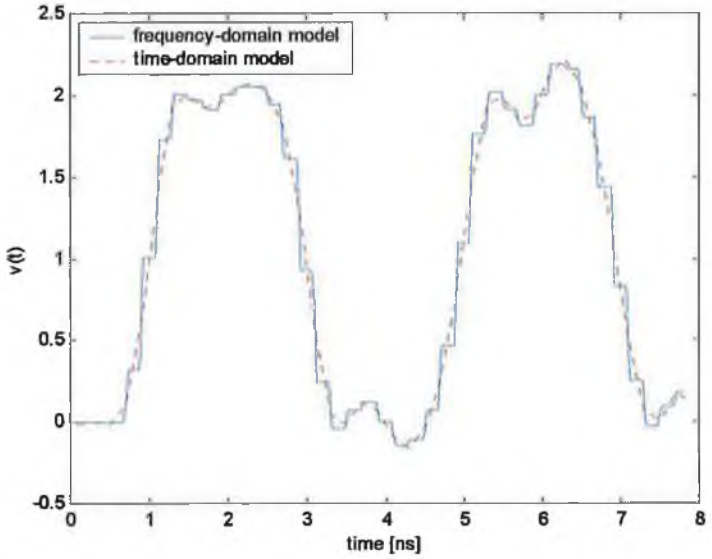


Fig. 5.5. *Step response for lossless low-pass filter system in Fig. 5.4.*

The time-domain result for the *s*-parameters model (frequency domain) of the network is obtained by numerical inversion of the Laplace transform. The step-shaped nature of the open-circuit response is expected due to the lossless nature of transmission lines. The time-domain model result is calculated using the method described in previous sections. The impulse response for each scattering parameter is obtained using (5.30) with 128 time-domain sample values. A corresponding FIR representation of each the scattering parameters with 128 coefficients is formed. A $\mathcal{Z}$-domain input-output transfer function is converted to state-space format and the Laguerre reduction technique is implemented with K set equal to 20 (i.e. 84% reduction in the system size). The computing cost is reduced by a factor of 6 over the direct convolution method. As evidenced by the result in Fig. 5.5, the proposed strategy is highly effective at capturing the essential nature of the response. The nature of the response is captured for a much lower computational cost than use of direct convolution techniques in conjunction with the 128 time-domain samples of the impulse response. Obviously, accuracy and efficiency can be effectively traded off against each other using the new strategy (reducing K increases efficiency but with a corresponding reduction in accuracy). Fig. 5.6 shows the absolute error between the full model and the reduced model as a function of the time from application of the input. This confirms the exponential weighting when a Laguerre model reduction algorithm is employed.

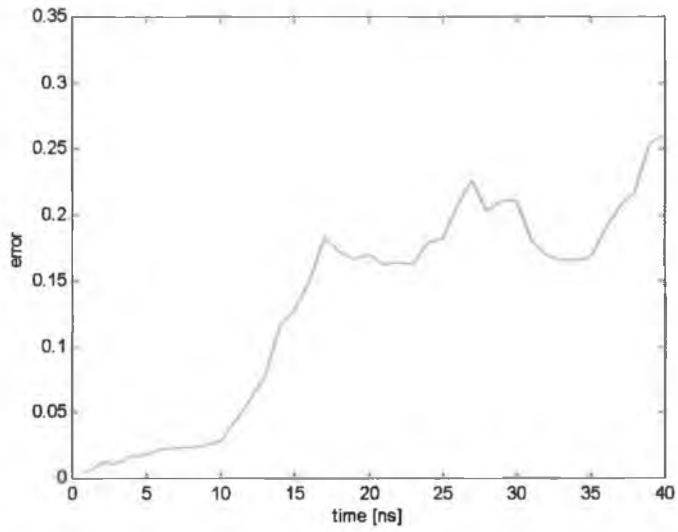


Fig. 5.6. *Absolute error between full model and reduced model.*

To further highlight the effectiveness of the proposed strategy, frequency-dependent losses are introduced into the same system. The losses are skin-effect losses in the transmission lines with a $\sqrt{\omega}$ frequency dependence

$$R = R_{dc} + R_s\sqrt{\omega}\,, \tag{5.50}$$

where R_{dc} = 0.3691 and R_s = 0.0126. Fig. 5.7 shows the frequency-domain result obtained from numerical inversion of the Laplace transform with a very high bandwidth. Superimposed on this is the result with the proposed interconnect modelling strategy. The reduced-order model once again has K set equal to 20.

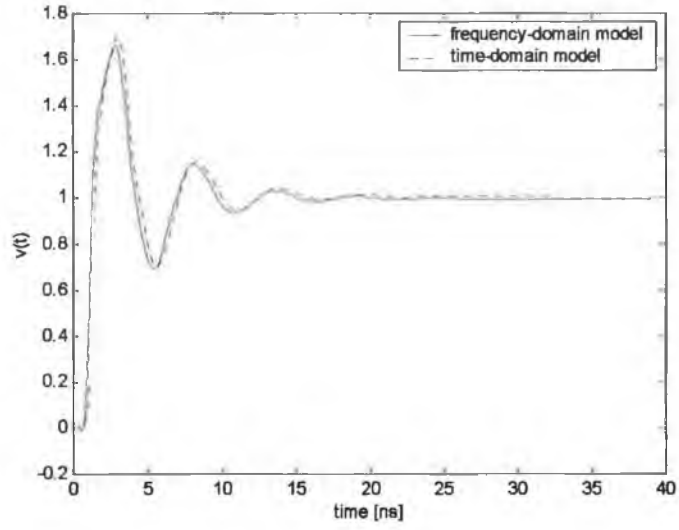


Fig. 5.7. *Result from low-pass filter structure inclusive of skin-effect losses.*

As before, the results are seen to indicate a high degree of accuracy for a low computational cost.

5.5.2. Ilustrative example 2 – The measured data

The next example involves the network topology shown in Fig. 5.8. It comprises the parallel connection of an open-circuit lossy transmission-line network and a 50Ω resistor all in series with a 50Ω source and lossy feeder transmission line. Measured scattering parameters are available for frequencies between 0 and 16GHz.

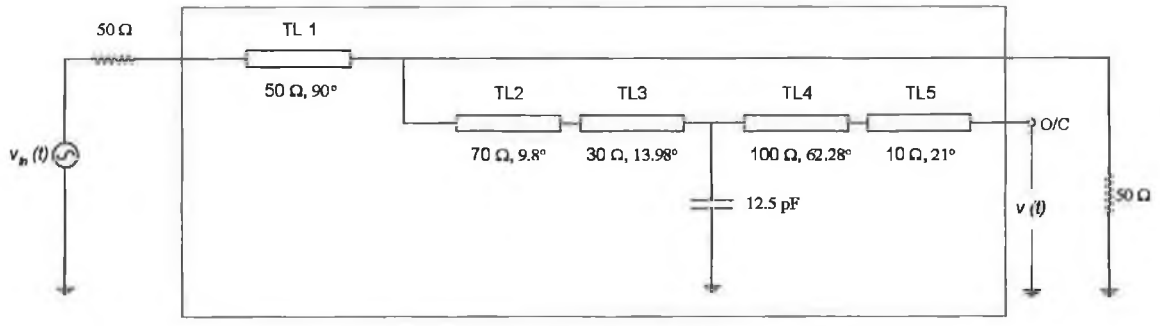
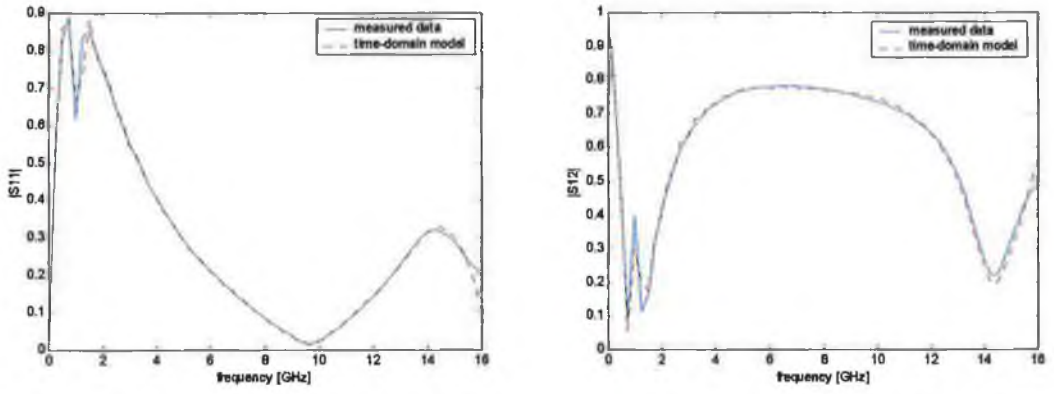


Fig. 5.8. *Linear interconnect network*

Fig. 5.9 shows the magnitude of the measured scattering parameters for S_{11} and S_{12}. The Hilbert transform relationship is applied to the data to ensure causality and the system is modelled as described in previous sections. The dashed lines in Fig. 5.9 show the scattering parameters resulting from the developed macromodel with K=20.



a) |S11| *b)|S12|*

Fig. 5.9. *Magnitude of measured responses of S_{11} and S_{12}*

Fig. 5.10 shows the result for a trapezoidal input with an amplitude equal to 5V, a rise/fall-time equal to 200 ps and a total duration of 1.4 ns.

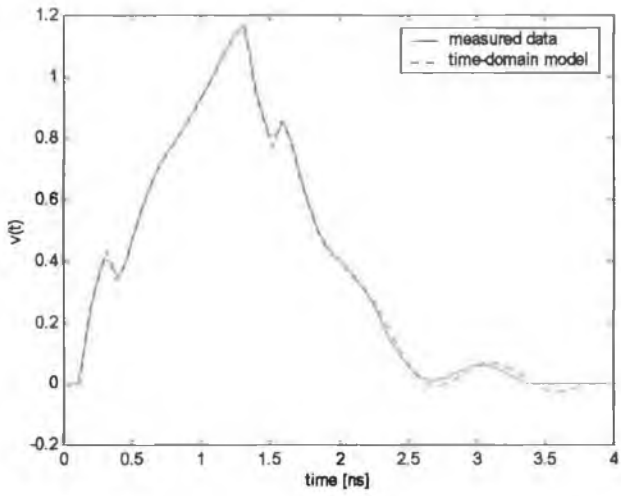


Fig. 5.10. *Pulse response from the circuit in Fig. 5.8.*

As can be seen from Fig. 5.10, excellent agreement between the measured result and the proposed time-domain model result is achieved.

5.6. Conclusions

An efficient and accurate modelling strategy for non-uniform interconnect networks characterised by frequency-domain *s*-parameters is presented in this chapter. The parameters may be obtained either from measurements or rigorous full-wave simulation. The method is especially suitable for interconnect networks for which analytical models cannot be obtained due to the complexity and inhomogeneity of the geometries involved. The proposed method can readily be implemented both for lossless/lossy non-uniform interconnects with frequency-dependent parameters.

Furthermore, the efficiency of the proposed technique is improved by employing a Laguerre-based model order reduction strategy to reduce the order of the discrete approximation of the system. The final model achieves both high efficiency and accuracy.

CHAPTER 6

Numerical Algorithms for the Transient Analysis of High Frequency Non-Linear Circuits

In general, simulation of modern integrated circuits requires at some point that a numerical solution be found for a system of typically highly *non-linear* stiff ordinary differential equations. Usually these equations arise from non-linear equivalent circuit models for discrete active devices. However, the complexity of modern circuits is such that the equations are highly stiff resulting in unacceptably long simulations. Hence, in order to cope with the complexity of modern integrated circuits, new numerical algorithms that take into account the nature of the differential equations arising in the transient analysis of non-linear circuits are needed.

The aim of this chapter is to suggest new numerical algorithms that may be utilised in modern circuit simulators. In total, four new numerical methods for solving an initial value problem (IVP) are proposed in this contribution. The new methods are compared to the widely used Adams-Moulton method to confirm their accuracy and efficiency.

6.1. Introduction

The "core" of a CAD tool is the discrete time numerical integrator that is required to solve the variety of non-linear differential equations that arise from mathematical models that describe circuit behaviour. Today these engines are struggling to cope with the complexity of the circuits that need to be simulated. The nature of modern circuit models used is such that '*stiff*' differential equations govern their behaviour. This is due to the very short time constants arising from internal charge dynamics and efforts to describe non-quasistatic behaviour. Very often it is required to perform an RF simulation of a circuit using digital modulation formats with long bit sequences. This results in an extended solution period and excessively slow computation.

Most existing general-purpose circuit simulators use implicit numerical integration techniques with adaptive time stepping. These methods have good stability

and accuracy control properties, but require the solution of non-linear algebraic equations at each time-step, which is computationally expensive. In addition, predictor-corrector methods can be used. In the traditional predictor-corrector setup, an explicit method uses polynomial extrapolation to provide an estimate of the solution at the next time step. The estimate is subsequently corrected using one iteration of an implicit formula. This usually results in a more accurate solution at a given time-point and improved stability of the integration scheme. In relation to stiff problems, the maximum allowable time-step that can be used with the traditional predictor-corrector techniques such as the Adams-Moulton technique may be unacceptably small. Hence, there is a need for specialised numerical techniques that enable utilisation of larger time-step during the solution process.

In the following section, a short survey of the existing numerical methods for the solution of initial value problems is given. In addition, some basic notations and concepts related to ODEs are introduced.

6.2. A short survey of numerical methods for the solution of initial value problems (IVP)

The initial value problem (IVP) arises in many areas of scientific research. It is rarely possible to find an analytical form of the solution to this problem; instead a *numerical approximation* to the true solution is the only possible approach. This approximation is also termed the *numerical solution* of the IVP. There are vast number of published papers and books dealing with the problem of finding a numerical solution for the IVP. The following short survey is based on some excellent references [J84], [HNW87], [B00], [H62], [G71], [AP98] but this is not, by any means, an exhaustive list of available sources.

6.2.1. Formulation of the IVP

An *ordinary differential equation (ODE)* of the *first order* is an equation of the form:

$$y' = \frac{dy}{dt} = f(t, y) \tag{6.1}$$

where the scalar function $f(t,y)$ is known. A function $y(t)$ is called a *solution* of equation (6.1) if for all t following equation is satisfied:

$$y'(t) = f(t, y(t)). \tag{6.2}$$

Newton, Leibniz and Euler observed that such a solution usually contains a free parameter, in other words, that it is possible to find many $y(t)$, which satisfy equation (6.2). But if the condition (*initial value condition*):

$$y(t_0) = y_0 \tag{6.3}$$

is imposed on the required solution $y(t)$, then the solution to the ODE is *uniquely* determined (Cauchy (1824)), i.e. there is only one function $y(t)$ which satisfies the following:

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0 . \tag{6.4}$$

Equation (6.4) represents an *initial value problem (IVP)* in the overall category of *ordinary differential equations (ODE)*. In the early investigation of IVP, y and f were regarded as scalar value functions but there is no reason not to extend this representation to a system of n ordinary differential equations (6.5)

$$\begin{aligned} &y_1' = f_1(t, y_1, \ldots, y_n) \ , \ y_1(t_0) = y_{10} \\ &\ldots \\ &y_n' = f_n(t, y_1, \ldots, y_n) \ , \ y_n(t_0) = y_{n0}. \end{aligned} \tag{6.5}$$

Peano (1890) introduced vector notation:

$$\boldsymbol{y} = (y_1, \ldots, y_n)^T, \qquad \boldsymbol{f} = (f_1, \ldots f_n)^T \tag{6.6}$$

that enables (6.5) to be written in form of (6.4) but this time, $\boldsymbol{y}$ and $\boldsymbol{f}$ are considered as *vector* valued functions.

Today, it is quite common to consider *autonomous* systems of differential equations given in the following form:

$$\boldsymbol{y}'(t) = \boldsymbol{f}(\boldsymbol{y}(t)) \tag{6.7}$$

because, should it be necessary, t can always be added to vector $\boldsymbol{y}(t)$ as an additional component which satisfies the trivial differential equation

$$\frac{dt}{dt} = 1 . \tag{6.8}$$

Finding the solution to the initial value problem has proven to be a very significant issue in many areas of mathematics, science and technology. Very often IVPs for partial differential equations (PDE) can conveniently be transformed into a system of ODEs, for example, with finite differences or finite element approximations in the variable x. Many problems in physics at the end of the day require solving one or more differential equations in order to find the required behaviour of observed system. When mathematical models of real electronic systems are used in order to simulate their dynamic behaviour, it inevitably results in a system of ODEs that needs to be solved. But

before an attempt to find a numerical solution of an IVP is made, it is necessary to determine if a solution exists, and if it does, how many solutions can be found for given problem. The following, well-known theorem summarises the conditions on the existence and uniqueness of the solution to IVPs [J84].

THEOREM 6.1. *(about existence and uniqueness of the solution to an IVP)*
Assume that the function f in (6.1) satisfies the following conditions:

i. *f(t,y) is a real function,*
ii. *f(t,y) is defined and continuous in the strip* $t \in [a,b]$, $y \in (-\infty,+\infty)$,
iii. *there exists a constant L such that for any* $t \in [a,b]$ *and for any two numbers* y_1 *and* y_2 *the following statement is satisfied*

$$|f(t,y_1) - f(t,y_2)| \leq L\,|y_1 - y_2| \tag{6.9}$$

where L is called the Lipschitz constant. Then, for any y_0 *the IVP (6.4) has a unique solution y(t) for* $t \in [a,b]$.

Proof of Theorem 6.1. can be found in a variety of books dealing with differential equations, e.g. [H62], and one interesting formulation and proof is given in [HNW87] as well as further investigation into the problem of the existence of solutions to ODEs and related areas.

Now assuming that the conditions in Theorem 6.1. are satisfied, i.e. there exists a unique solution for (6.4), next step is to find a methodology for obtaining the solution. In a very small number of cases, it is possible to find the solution for (6.4) in an analytical form. More often, it is only possible to find a *numerical approximation* to the true solution (6.2). There are two basic approaches [G71] for obtaining numerical approximation. One is to represent an approximate solution by the sum of a finite number of independent functions, i.e. a truncated power series or the first few terms of an expansion in orthogonal functions. This approach is considered of more theoretical value as inclusion of such a representation in an algorithm that can be executed on a digital computer presents a great challenge. The second approach is the *difference method* or *step-by-step method*, which provides a rule for computing the approximation at step n to $y(t_n)$ in terms of values of y at t_{n-1} and other preceding points. These methods are generally more suited for automatic computations. Hence, the short survey presented in the following section will be limited only to step-by-step methods.

6.2.2. Elements of numerical methods for solving IVP

The numerical step-by-step methods for the solution of (6.4) are algorithms which produce a table of approximate values to $y(t)$ at certain equally spaced points called *grid, nodal, net* or *mesh points* along the t coordinate. Each grid point is given by the relationship:

$$\begin{aligned} t_{n+1} &= t_n + h,\ n = 0, 1, 2, \ldots, N-1 \\ t_0 &= a,\ t_N = b \end{aligned}, \tag{6.10}$$

where h is called the *step size* and $[a, b]$ is the *interval* on which the required solution is sought. Sometimes it is useful to write (6.10) in following form:

$$t_n = t_0 + nh, \quad n = 1, \ldots, N. \tag{6.11}$$

When using numerical methods to find the solution of an IVP, what is involved is a calculation of an approximation y_n to a solution $y(t)$ for $t = t_n$. These approximate values of $y(t)$ usually contain errors, namely *round-off error* and/or *truncation error*. The round-off errors are caused by the finite computer representation of a number. The truncation error is caused by the numerical method itself (e.g. taking a finite number of terms in the Taylor series expansion) and has nothing to do with the computer properties.

DEFINITION 6.1. *(Round-off error)*

The *round-off error* is the quantity R that must be added to a finite representation of a computed number in order to make it equal to the representation of number that the numerical algorithm would give if the computer had infinite precision, i.e.

y(machine representation) + R = y(representation)

DEFINITION 6.2. *(Truncation error)*

The *truncation error* is the quantity T that must be added to the representation of the computed quantity in order to make the result exactly equal to the quantity that is sought, i.e.

y(representation) + T = y(exact)

Two issues that are very important in relation to the numerical solution of IVPs are *stability and convergence*.

DEFINITION 6.3. *(Convergence)*

The concept of *convergence* refers to the fact that any desired degree of accuracy can be achieved for any problem satisfying the Lipschitz condition (6.9) by picking a suitably small step size h.

DEFINITION 6.4. *(Stability)*

If there exists an h_0 for each differential equation such that a change in the initial value by a fixed amount produces a bounded change in the numerical solution for all $0 \le h \le h_0$, then the method is deemed *stable*.

These definitions [J84] are very loosely given – the intention here is to present concepts.

6.2.3. Numerical methods for solving IVP

The numerical methods for finding the solution to the initial value problem are usually classified into two types [J84]:

i. *Singlestep Methods* – These methods enable an approximation to the true solution $y(t)$ at t_{n+1} to be found, if y_n, y_n' and h are known.

ii. *Multistep Methods* – These methods use recurrence relations, which express the value of $y(t)$ at t_{n+1} in terms of the values of $y(t)$ and derivative values $y'(t)$ at t_n and at previous nodal points.

In addition, there is a whole range of existing numerical methods that cannot be classified as either of the above: Taylor series methods, Hybrid methods, Cyclic composite methods, Rosenbrock methods, etc. A good introduction to these methods can be found in [B00].

6.2.3.1. Singlestep Methods

The Taylor Series Expansion about the point t or t_n gives the basis for most one-step numerical integration formulas:

$$y(t+h) = y(t) + hy'(t) + \frac{h^2}{2!}y''(t) + \frac{h^3}{3!}y'''(t) + \ldots \tag{6.12}$$

where $h = t_{n+1} - t_n$. Since $y' = f(t,y)$, $y'' = f'(t,y)$,..., the equation (6.12) becomes:

$$y(t+h) = y(t) + hf(t,y) + \frac{h^2}{2!}f'(t,y) + \frac{h^3}{3!}f''(t,y) + \ldots \tag{6.13}$$

Letting $t \to t_n$ and considering that, (6.13) can be written in discrete notation as:

$$y_{n+1} = y_n + hf(t_n, y_n) + \frac{h^2}{2!} f'(t_n, y_n) + \frac{h^3}{3!} f''(t_n, y_n) + \ldots \tag{6.14}$$

where $y_i = y(t)$, $i=0,1,\ldots$ Therefore, a general singlestep method can be written in the form:

$$y_{n+1} = y_n + h\phi(t_n, y_n, h), n = 0,1,\ldots,N-1, \tag{6.15}$$

where ϕ is function of the arguments t, y, h and, in addition, depends on the right-hand side of (6.4). The function $\phi(t,y,h)$ is called an *increment function*. If y_{n+1} can be obtained simply by evaluating the right-hand side of (6.4), then the singlestep method is termed *explicit* otherwise it is termed *implicit*. The most common singlestep methods are the Euler's method and the Runge-Kutta family of methods.

Truncation of the series expansion in (6.14) introduces a truncation error. The exact value $y(t)$ will satisfy:

$$y(t_{n+1}) = y(t_n) + h\phi(t_n, y_n, h) + T_n, \quad n = 0,1,\ldots,N-1, \tag{6.16}$$

where T_n is the truncation error. The largest integer p such that $\left|h^{-1} T_n\right| = O(h^p)$ is called the *order* of the singlestep method.

Forward and Backward Euler Method

The simplest singlestep method is the *Forward Euler (FE)* method. It truncates the Taylor series after the 1st order term, giving

$$y_{n+1} = y_n + hf(t_n, y_n). \tag{6.17}$$

The Euler's method is a 1st order explicit method whose truncation error per step is of the order $O(h^2)$.

The implicit version of Eulers' method is known as the *Backward Euler (BE) method.* It is derived in the same manner as the FE method, except that everything is centred around t_{n+1} rather than t_n, yielding following implicit formula:

$$y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}) \tag{6.18}$$

Geometrically, instead of using the tangent at (t_n, y_n), as in Forward Euler method, the Backward Euler method uses the tangent at the future point (t_{n+1}, y_{n+1}), thus enhancing the stability of method which proves to be very useful when dealing with so called *stiff* problems characteristic of electronic circuit models. But, of course, there is a price to be paid for enhanced stability: while the FE method is explicit, the BE method is implicit. It means that the unknown variable y_{n+1} at each step appears on both sides of equation

(6.18) which is generally a non-linear expression. Consequently, a non-linear system of algebraic equations has to be (approximately) solved at each step, which can be computationally prohibitive for large systems.

Trapezoidal Method

The derivation of the Euler methods is based on a Taylor expansion centred at t_n for the Forward Euler, and at t_{n+1} for the Backward Euler. The next logical step is to attempt to form an expansion in the middle of the interval $[t_n, t_{n+1}]$, i.e. around

$$t_{n+\frac{1}{2}} = t_{n+1} - \frac{1}{2}h. \tag{6.19}$$

After some mathematical calculations, the following formula is obtained:

$$\frac{y(t_{n+1}) - y(t_n)}{h} = \frac{1}{2}(y'(t_{n+1}) + y'(t_n)) - \frac{h^2}{12}y'''(t_{n+1/2}) + O(h^4). \tag{6.20}$$

Disregarding the parts with higher derivatives gives the following formula for the *trapezoidal (TR) method:*

$$y_{n+1} = y_n + \frac{h}{2}(f(t_{n+1}, y_{n+1}) + f(t_n, y_n)) \tag{6.21}$$

This method is more accurate (second-order accurate) than Euler's and it is implicit (like Backward Euler).

Runge-Kutta Methods

One important group of singlestep methods are the *Runge-Kutta (RK) methods.* These methods refer to a whole range of methods that use a truncated Taylor series expansion without requiring the calculation of the higher derivatives. Consider the Mean-Value Theorem which states that any solution of (6.4) satisfies:

$$y(t_{n+1}) = y(t_n) + hy'(\xi_n) = y(t_n) + hf(\xi_n, y(\xi_n)), \tag{6.22}$$

where $\xi_n = t_n + \theta_n h$, $0 < \theta_n < 1$. Setting $\theta_n = 1/2$ and using of Euler's method with spacing $h/2$, yields:

$$y(t_n + \frac{h}{2}) \cong y_n + \frac{h}{2}f(t_n, y_n). \tag{6.23}$$

Thus, the following approximation is obtained:

$$y_{n+1} = y_n + hf(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n)). \tag{6.24}$$

Alternatively, and again using Euler's method, it is possible to write following:

$$y'(t_n+\frac{h}{2}) \cong \frac{1}{2}[y'(t_n)+y'(t_{n+1})] \cong \frac{1}{2}[f(t_n,y_n)+f(t_{n+1},y_n+hf_n)], \tag{6.25}$$

yielding the approximation:

$$y_{n+1} = y_n + \frac{h}{2}[f(t_n,y_n)+f(t_{n+1},y_n+hf(t_n,y_n))]. \tag{6.26}$$

Either (6.24) or (6.26) can be regarded as

$$y_{n+1} = y_n + h(average\ slope). \tag{6.27}$$

This is the underlying idea of the Runge-Kutta approach: find the slope at t_n and at several other points, average these slopes, multiply by h, and add the result to y_n yielding the following *end-of-step* value:

$$y_{n+1} = y_n + \Phi(x_n, y_n, h), \tag{6.28}$$

where $\Phi(x_n, y_n, h)$ is the increment function whose general form is

$$\Phi(x_n, y_n, h) = \sum_{i=1}^{N} w_i K_i . \tag{6.29}$$

N is the number of stages in the RK method, w_i are arbitrary parameters and the explicit formulae for the K_i are determined by comparison with the original Taylor series expansion with appropriate approximations for the derivatives of $f(t_n,y_n)$ thus yielding:

$$K_i = hf(t_n + c_i h,\ y_n + \sum_{j=1}^{i-1} a_{ij} K_j)\ ,\ c_1 = 0,\ i = 1,2,...,\ N\ , \tag{6.30}$$

where parameters c_2, c_3, ..., c_N, a_{2j}, ... $a_{N(N-1)}$ are arbitrary. The increment function can be interpreted as the linear combination of the slopes at t_n and at several other points between t_n and t_{n+1}. To obtain specific values for the parameters, y_{n+1} is expanded in powers of h such that it agrees with the Taylor series expansion of the solution of the differential equation to a specified number of terms. For example, the classical 4th order RK method has following parameters:

$$\begin{aligned} K_1 &= hf(x_n, y_n) \\ K_2 &= hf(x_n + \frac{1}{2}h,\ y_n + \frac{1}{2}K_1) \\ K_3 &= hf(x_n + \frac{1}{2}h,\ y_n + \frac{1}{2}K_2) \\ K_4 &= hf(x_n + h,\ y_n + K_3) \end{aligned} \tag{6.31}$$

and

$$y_{n+1} = y_n + (\frac{1}{6}K_1 + \frac{1}{3}K_2 + \frac{1}{3}K_3 + \frac{1}{6}K_4) \tag{6.32}$$

RK methods are very well known and widely used in their implicit and explicit forms and for numerous applications [J84]. Probably, the most famous and widely used formula is, the so called, *Runge-Kutta-Fehlberg* (RKF) pair which has six stages and is a method of order 4 with an error estimate or a method of order 5 without an error estimate. The parameters for the RKF are:

$$\begin{aligned}
K_1 &= hf(x_n, y_n) \\
K_2 &= hf(x_n + \frac{1}{4}h, y_n + \frac{1}{4}K_1) \\
K_3 &= hf(x_n + \frac{3}{8}h, y_n + \frac{3}{32}K_1 + \frac{9}{32}K_2) \\
K_4 &= hf(x_n + \frac{12}{13}h, y_n + \frac{1932}{2197}K_1 - \frac{7200}{2197}K_2 + \frac{7296}{2197}K_3) \\
K_5 &= hf(x_n + h, y_n + \frac{439}{216}K_1 - 8K_2 + \frac{3680}{513}K_3 - \frac{845}{4104}K_4) \\
K_6 &= hf(x_n + \frac{1}{2}h, y_n - \frac{8}{27}K_1 + 2K_2 - \frac{3544}{2565}K_3 + \frac{1859}{4104}K_4 - \frac{11}{40}K_5)
\end{aligned} \tag{6.33}$$

and

$$y_{n+1} = y_n + (\frac{16}{135}K_1 + \frac{6656}{12825}K_3 + \frac{28561}{56430}K_4 - \frac{9}{50}K_5 + \frac{2}{55}K_6) \tag{6.34}$$

$$\hat{y}_{n+1} = y_n + (\frac{25}{216}K_1 + \frac{1408}{2565}K_3 + \frac{2197}{4104}K_4 - \frac{1}{5}K_5) \tag{6.35}$$

where $\hat{y}_{n+1}$ is used for error estimation [AP98].

There are many variations of the above formulae but for this overview only the most widely used are presented. More details on the techniques presented here as well as others can be found in [J84], [HNW87], [B00], [H62], [G71], [AP98].

6.2.3.2. Multistep methods

In multistep methods, an estimation of the solution at the next time step is first obtained using an explicit method utilising polynomial extrapolation. Then, the estimate is corrected using one iteration of an initial function. The general multistep method may be written in the form:

$$y_{n+1} = a_1 y_n + a_2 y_{n-1} + ... + a_k y_{n-k+1} + h\phi(t_{n+1}, t_n, ..., t_{n-k+1}; y'_{n+1}, y'_n, ..., y'_{n-k+1}; h) \tag{6.36}$$

where h is a constant stepsize and a_1, a_2, ... a_k are real constants. If ϕ is independent of y'_{n+1}, then the general multistep method is called an *explicit*, *open* or *predictor method.*

Otherwise, it is called an *implicit, closed* or *corrector method.* The true value $y(t_{n+1})$ will satisfy:

$$y(t_{n+1}) = a_1 y(t_n) + \ldots + a_k y(t_{n-k+1}) + h\phi(t_{n+1}, t_n, \ldots, t_{n-k+1}; y'_{n+1}, y'_n, \ldots, y'_{n-k+1}) + T_n(y(t_n), h), \quad n = 0,1,\ldots,N-1 \tag{6.37}$$

where T_n is the truncation error. The largest integer p such that $| h^{-1} T_n(y(t_n),h) | = O(h^p)$ is the *order* of general multistep method.

One very important class of multistep methods is the general *linear multistep method* given by a following linear form:

$$y_{n+1} = a_1 y_n + a_2 y_{n-1} + \ldots + a_k y_{n-k+1} + h(b_0 y'_{n+1} + b_1 y'_n + \ldots + b_k y'_{n-k+1}). \tag{6.38}$$

The constants a_i and b_i are real and known. The k-1 values $y_1, y_2, \ldots, y_{k-1}$ required to start the computation in (6.38) are obtained using the single step methods. One common notation for (6.38) for linear multistep methods is given by (6.39)

$$\sum_{j=0}^{k} \alpha_j y_{n-j} = h \sum_{j=0}^{k} \beta_j f_{n-j}, \tag{6.39}$$

where α_j, β_j are the coefficients. For practical purposes, usually it is assumed that $\alpha_0 \neq 0$ and $|\alpha_k| + |\beta_k| \neq 0$. To eliminate arbitrary scaling, α_0 is set equal to 1. Obviously, for the explicit linear multistep method $\beta_0 = 0$ and for the implicit method $\beta_0 \neq 0$.

Linear multistep methods usually come in families. The most popular are the *Adams family* and the *Backward Differentiation Formula (BDF) family.* The common feature of most linear multistep methods is that they are based on polynomial interpolation.

Adams methods

Given the initial value problem (6.4), it is possible to integrate over a finite step using an interpolating polynomial, which passes through previously computed values of $f(t, y(t))$ that are within the interval, as in

$$y_n = y_{n-1} + \int_{t_{n-1}}^{t_n} f[t, y(t)]dt. \tag{6.40}$$

For all Adams methods, the relevant coefficients in (6.39) are set as follows:

$$\alpha_0 = 1, \ \alpha_1 = -1 \text{ and } \alpha_j = 0, \ j > 1$$

The k-step *explicit Adams method* (also called the *Adams-Bashforth (AB) method* [BA883]) is obtained by interpolating f through the k previous points: $t_{n-1}, t_{n-2}, \ldots, t_{n-k}$, which yields the following formula:

$$y_n = y_{n-1} + h\sum_{j=1}^{k} \beta_j f_{n-j}\,, \tag{6.41}$$

where

$$\beta_j = (-1)^{j-1} \sum_{i=j-1}^{k-1} \binom{i}{j-1} \gamma_i \tag{6.42}$$

and

$$\gamma_i = (-1)^i \int_0^1 \binom{-s}{i} ds\,. \tag{6.43}$$

The *k*-step Adams-Bashforth methods are of order $p=k$ and they are explicit methods with very small regions of absolute stability. This has inspired research into implicit versions of the Adams methods.

The *k*-step *implicit Adams method* (also called the *Adams-Moulton (AM) method*) is derived in a similar manner to the explicit method but the interpolating polynomial interpolates f at the unknown value at t_n as well, which yields following formula:

$$y_n = y_{n-1} + h\sum_{j=0}^{k} \beta_j f_{n-j} \tag{6.44}$$

A straightforward use of interpolation yields the appropriate coefficients that are given in literature dealing with Adams methods, e.g.[AP98]. The *k*-step Adams-Bashforth methods are of order $p = k+1$ which follows immediately from the fact that $k+1$ points are used in the underlying polynomial interpolation. An exception is in the case for $k = 1$, where f_{n-1} is not used, yielding $p = k = 1$.

The Adams-Moulton methods have smaller error constants than the Adams-Bashforth methods of the same order and use one fewer step for the same order. They have much larger stability regions than the Adams-Bashforth methods. But since Adams-Moulton methods are implicit, their implementation is not so straightforward so they are often used together with Adams-Bashforth methods for the solution of ODEs in a form of implementation known as a *predictor-corrector*. In this type of implementation, an approximation y_n^0 to y_n is predicted, usually by an explicit multistep method of the same order as the implicit method

$$\mathrm{P}:\ y_n^0 + \hat{\alpha}_1 y_{n-1} + \ldots + \hat{\alpha}_k y_{n-k} = h(\hat{\beta}_1 f_{n-1} + \ldots + \hat{\beta}_k f_{n-k}). \tag{6.45}$$

Then the function is *evaluated* at y_n^0:

$$\mathrm{E}:\ f_n^0 = f(t_n, y_n^0) \tag{6.46}$$

and inserted into the corrector formula to obtain a new approximation to y_n. Setting the iteration count $\nu = 0$ yields in general

$$\mathrm{C}: \quad y_n^{\nu+1} + \alpha_1 y_{n-1} + \ldots + \alpha_k y_{n-k} = h(\beta_0 f_n^{\nu} + \beta_1 f_{n-1} + \ldots + \beta_k f_{n-k}). \tag{6.47}$$

The procedure can be stopped here (this is called *PEC* method), or the function can be evaluated at y_n^1 to give

$$\mathrm{E}: \quad f_n^1 = f(t_n, y_n^1) \tag{6.48}$$

(this is called *PECE* method), or the steps E and C can be iterated ν times to form $P(EC)^{\nu}$ or a $P(EC)^{\nu}E$ method. The final function evaluation in a $P(EC)^{\nu}E$ method yields a better value for f to be used in next time step as the new f_{n-1}. Although it appears that the $P(EC)^{\nu}E$ method might be more expensive, the final function evaluation is usually advantageous because it yields a significant increase in the region of absolute stability when compared to the corresponding $P(EC)^{\nu}$ method.

The most widely used variant of predictor-corrector methods is the PECE. For example, consider the following algorithm based on the 4th order Adams formulae:

1. Since multistep methods are not self-starting one needs to use a singlestep method to get values for y_3, y_2, y_1 – e.g. utilising Runge-Kutta-Fehlberg of 4th order. y_0 is known since it is the initial condition.
2. Calculate y_{n+1}^* using the 4th order formula for the Adams-Bashforth method:

$$y_{n+1}^* = y_n + \frac{h}{24}(55 f_n - 59 f_{n-1} + 37 f_{n-2} - 9 f_{n-3})$$

3. Evaluate $f_{n+1}^* = f(t_{n+1}, y_{n+1}^*)$
4. Calculate y_{n+1} using the 4th order formula for Adams-Moulton method

$$y_{n+1} = y_n + \frac{h}{24}(9 f_{n+1}^* + 19 f_n - 5 f_{n-1} + f_{n-2})$$

5. Increment $t_{n+1} = t_n + h$, go to step 2 and repeat until finished.

This variant of the Adams-Moulton method is implemented and used for comparative purposes against the novel methods proposed in this dissertation.

Backward differentiation formula (BDF) methods

Another set of popular multistep methods for stiff problems are the Backward Differentiation Formula (BDF). Their distinguishing feature is that $f(t, y)$ is evaluated

only at the right end of the current step, (t_n, y_n) (t_n, y_n). A motivation behind this is to obtain formulae with the fast decay property. This is done by setting $\beta_0 \neq 0$ and $\beta_j = 0, j > 0$ in (6.39). In contrast to the Adams methods, which were derived by *integrating* the polynomial which interpolates past values of f, the BDF methods are derived by *differentiating* the polynomial which interpolates past values of y and setting the derivative at t_n to $f(t_n, y_n)$. This yields the k-step BDF which has order $p = k$ and can be written in the form:

$$\sum_{i=0}^{k} \alpha_i y_{n-i} = h\beta_0 f(t_n, y_n) \tag{6.49}$$

where $\alpha_0 = 1$. The BDF methods are implicit and are usually implemented in conjunction with a modified Newton method to solve the non-linear system at each step. However, this is time-consuming and may be computationally prohibitive in the case of large systems.

6.3. The problem of stiffness

Stiff ordinary differential equations arise in many areas of electronic circuit analysis and simulation. Most CAD techniques that have been developed for circuit simulation suffer from the problem and inefficiency when simulating complex electronic circuits described with stiff ODEs. In order to illustrate the problem of stiffness, it is necessary to revert to the very beginning of the problem - finding a numerical solution for an IVP. The first and foremost requirement for this task is to make the difference between the true and the calculated solution as small as possible, i.e. to ensure that the obtained solution is *accurate*. In order to satisfy this accuracy requirement, the stepsize h must be chosen such that it is deemed sufficient. This usually involves some form of error estimation. Ideally, the choice of stepsize h should be dictated only by the approximation accuracy requirement. But it turns out, that for many of the numerical methods in use (e.g. Euler, Runge-Kutta, Adams methods), h must be chosen sufficiently small to obey an additional, absolute stability restriction, as well. Loosely speaking, the IVP is referred to as being *stiff* if this absolute stability requirement dictates a much smaller stepsize h than is needed to satisfy the accuracy requirements alone [AP98]. Ascher and Petzold [AP98] define stiffness in terms of the behaviour of an explicit difference method, e.g. forward Euler as:

DEFINITION 6.5. *(Stiffness)*

An IVP is *stiff* in a given interval of integration if the step size needed to maintain the stability of the forward Euler method is much smaller than the step size required to represent the solution accurately.

It should be noted that, in addition to the differential equation itself, stiffness depends on the accuracy criterion imposed, the length of the interval of integration and the region of absolute stability of the method used.

The phenomenon of stiffness is usually found in systems incorporating behaviours with greatly differing time constants. The *time constant* is the term used by engineers and physicists to refer to the rate of decay of a response. For example, the equation

$$y' = \lambda y \tag{6.50}$$

has the solution $ce^{\lambda t}$. If λ is negative, then y decays by a factor e^{-1} in time $-1/\lambda$. This term is called the time constant, τ. Physical systems frequently behave, at least locally, in an exponential fashion, e.g. capacitors discharging. In a complex electronic circuit, different components will be decaying at different rates. For the system described by:

$$\boldsymbol{y}' = \boldsymbol{f}(y) \tag{6.51}$$

the decay rates may be related locally to the eigenvalues of $\partial \boldsymbol{f} / \partial \boldsymbol{y}$. If some of the components are slow and others are fast, the fast ones will control the stability of the method, although the components may have decayed to insignificant levels. For example, consider the following system:

$$\begin{aligned} y' &= -y, & y(0) &= 1 \\ z' &= -100z, & z(0) &= 1 \end{aligned} \tag{6.52}$$

These equations are independent of each other, so it is possible to analyze the behaviour of each one separately. For the most of the numerical methods in use, the stability requirements will necessitate the step size h to be smaller than 1/100 [G71]. Hence, the integration step for equations (6.52) is stipulated by the time constant of z. However, after a few steps, the value of z will be so small that it will be negligible compared to y as can be seen in Fig 6.1. Nevertheless very small steps must be used because of the second component z, although only the first component y contains any significant information. This illustrates the problem related to obtaining a numerical solution to stiff differential equations.

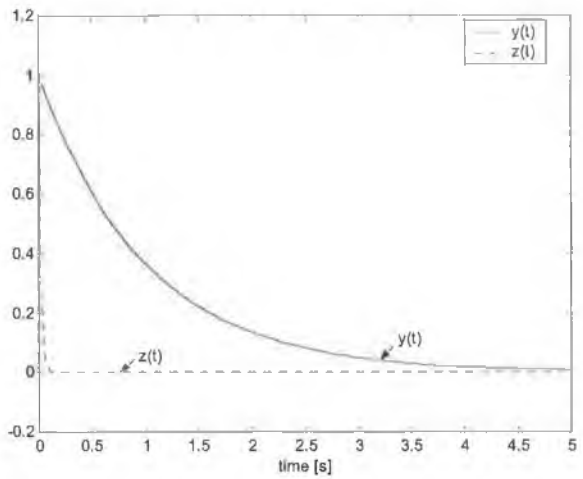


Fig. 6.1. *Illustration of stiffness problem*

One remedy to the problem given with (6.52) may involve separating (decoupling) the two components and using a different stepsize h or even different methods for each. However, in the general case, this separation of equations is not possible. For example, consider system (6.85) given later in this chapter with its solution given by (6.86). As is obvious, the solution for both dependant variables contains both fast and slow components thus yielding a restriction on the choice of stepsize h.

It is not necessary to consider a *system of equations* to observe the problem of stiffness. The stiffness may arise in a single ODE as well. For example, consider the following ODE [G71]:

$$y' = \lambda(y - F(t)) + F'(t), \quad \lambda \ll 0 \tag{6.53}$$

The solution to (6.53) is given by:

$$y = (y_0 - F(0))e^{\lambda t} + F(t). \tag{6.54}$$

For $y_0 - F(0) \neq 0$, λt will soon be sufficiently negative that the first component will be insignificant compared to the second. If the error equation for (6.53) is examined using any of the widely used methods [G71], it is seen that the local truncation error is determined by h and a derivative of F when $\lambda t \ll 0$, whereas the stability is dependent on the value of $h\lambda$. Since λ is a fixed parameter, h determines the stability. Therefore, for any smooth, slowly varying function $F(t)$, equation (6.53) has similar behaviour to the stiff system in (6.52). While it is true that numerical approximation of (6.53) by any one of the techniques discussed so far converges to the solution as $h \to 0$, h has to be intolerably small before acceptable accuracy is obtained in practice, so small in fact, that round-off errors and computation time become critical.

6.4. The proposed approach

From the discussion presented in Section 6.3, it is clear that special care has to be taken when solving potentially stiff ODEs, both from the accuracy and stability viewpoint. In addition, the efficiency of the ODE solver is the practical limiting factor for the performance of all circuit simulation techniques. Therefore, there is a need for a novel numerical algorithm that enables use of a longer timestep, thus improving the efficiency of the solver but retaining the required accuracy of the solution.

Consider the following initial value problem:

$$\frac{dy}{dx} = f(t, y(t)) \ , \ y(t_0) = y_0. \tag{6.55}$$

In order to solve (6.55), $y_i(t)$, is approximated by the Padé approximant given as:

$$y_i(t) = \phi_{m/n} = \frac{\sum_{j=0}^{m} a_j h^j}{\sum_{k=0}^{n} b_k h^k}, \quad h = t - t_{i-1}, \tag{6.56}$$

where $b_0 = 1$. The Padé approximation is chosen due to its excellent approximating properties [GW99]. A sequence of local approximations to $y(t)$ is then built in order to provide a solution to (6.55), in a manner similar to that proposed in [GN97]. The method is advanced in time by using the solution at time t as the initial condition for the next time step. The manner in which the coefficients of the Padé approximant ($a_j, j = 0,...,m$ and $b_k, k=1,..,n$) are obtained defines different methods for solving ODEs. These methods are presented in the reminder of this chapter.

6.5. Methods that do not use derivatives of the function $f(t, y(t))$

The two **new** methods that are presented in this section are named the Exact-fit and the Padé-fit method. The Exact-fit method is based on fitting dy/dt from (6.55) *exactly* over the past N points, in order to obtain the coefficients of a Padé approximant (6.56). The Padé-fit method utilises an initial polynomial approximation of dy/dt to calculate the coefficients of the Padé approximant from a set of linear equations.

6.5.1. Exact-fit method

The approach taken in the Exact-fit method is to fit dy/dt *exactly* over a number of past time points. For the purpose of clearer notation, the method will be presented for

a scalar IVP, with the note that its extension to a system of ODEs is straightforward. Without loss of generality, assume that the Padé approximant is chosen such that $m = n = 2$, i.e.

$$y(t) = \phi_{2/2} = \frac{a_0 + a_1 t + a_2 t^2}{1 + b_1 t + b_2 t^2}, \tag{6.57}$$

therefore

$$\frac{dy}{dt} = \frac{(a_1 - a_0 b_1) + 2(a_2 - a_0 b_2)t + (a_2 b_1 - a_1 b_2)t^2}{(1 + b_1 t + b_2 t^2)^2}. \tag{6.58}$$

As explained before, the current value of $y(t)$ is taken as the initial condition for next step ($t_i = t_0 = 0$) hence:

$$a_0 = y_i(t). \tag{6.59}$$

To obtain the remaining four coefficients, dy/dt is fitted *exactly* over the past 4 points by substituting t with t_0, t_0-h, t_0-2h, t_0-3h in (6.58). Subsequently, a 5x5 system of non-linear equations is obtained where a_0, a_1, a_2, b_1 and b_2 are unknowns, i.e.

$$\begin{aligned}
a_0 &= y_i(t) \\
F_1(a_1, a_2, b_1, b_2) &= a_1 - a_0 b_1 - f_0 \equiv 0 \\
F_2(a_1, a_2, b_1, b_2) &= a_1 - a_0 b_1 - 2(a_2 - a_0 b_2)h + (a_2 b_1 - a_1 b_2)h^2 - f_{-1}(1 - hb_1 + b_2 h^2)^2 \equiv 0 \\
F_3(a_1, a_2, b_1, b_2) &= a_1 - a_0 b_1 - 4(a_2 - a_0 b_2)h + 4(a_2 b_1 - a_1 b_2)h^2 - f_{-2}(1 - 2hb_1 + 4b_2 h^2)^2 \equiv 0 \\
F_4(a_1, a_2, b_1, b_2) &= a_1 - a_0 b_1 - 6(a_2 - a_0 b_2)h + 9(a_2 b_1 - a_1 b_2)h^2 - f_{-3}(1 - 3hb_1 + 9b_2 h^2)^2 \equiv 0
\end{aligned} \tag{6.60}$$

where

$$\begin{aligned}
f_0 &= f(t_0, y(t_0)), & f_{-2} &= f(t_0 - 2h, y(t_0 - 2h)), \\
f_{-1} &= f(t_0 - h, y(t_0 - h)), & f_{-3} &= f(t_0 - 3h, y(t_0 - 3h)).
\end{aligned}$$

Solving this system of nonlinear equations, e.g. by using the Newton method, for each step yields values for the coefficients a_0^i, a_1^i, a_2^i, b_1^i and b_2^i for that particular step. These are then used to calculate $y_{i+1}(t)$, $t = t_0 + h$ as:

$$y_{i+1} = \frac{a_0^i + a_1^i h + a_2^i h^2}{1 + b_1^i h + b_2^i h^2} \tag{6.61}$$

Implementation of a predictor-corrector algorithm is not complicated with this method. To find an expression for the corrector dy/dt is now fitted over t_0+h and the three previous points. This results in another 5x5 system of non-linear equations (6.62).

The coefficients $\tilde{a}_0$, $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{b}_1$ and $\tilde{b}_2$ are the new set of unknown coefficients that need to be calculated in order to obtain the value $y_{i+1}(t)$ in the corrector step, i.e.

$$\begin{aligned}
&\tilde{a}_0 = y_i(t)\\
&F_2(\tilde{a}_1, \tilde{a}_2, \tilde{b}_1, \tilde{b}_2) = \tilde{a}_1 - \tilde{a}_0\tilde{b}_1 - 2(\tilde{a}_2 - \tilde{a}_0\tilde{b}_2)h + (\tilde{a}_2\tilde{b}_1 - \tilde{a}_1\tilde{b}_2)h^2 - f_1(1 + h\tilde{b}_1 + \tilde{b}_2h^2)^2 \equiv 0\\
&F_2(\tilde{a}_1, \tilde{a}_2, \tilde{b}_1, \tilde{b}_2) = \tilde{a}_1 - \tilde{a}_0\tilde{b}_1 - f_0 \equiv 0\\
&F_3(\tilde{a}_1, \tilde{a}_2, \tilde{b}_1, \tilde{b}_2) = \tilde{a}_1 - \tilde{a}_0\tilde{b}_1 - 2(\tilde{a}_2 - \tilde{a}_0\tilde{b}_2)h + (\tilde{a}_2\tilde{b}_1 - \tilde{a}_1\tilde{b}_2)h^2 - f_{-1}(1 - h\tilde{b}_1 + \tilde{b}_2h^2)^2 \equiv 0\\
&F_4(\tilde{a}_1, \tilde{a}_2, \tilde{b}_1, \tilde{b}_2) = \tilde{a}_1 - \tilde{a}_0\tilde{b}_1 - 4(\tilde{a}_2 - \tilde{a}_0\tilde{b}_2)h + 4(\tilde{a}_2\tilde{b}_1 - \tilde{a}_1\tilde{b}_2)h^2 - f_{-2}(1 - 2h\tilde{b}_1 + 4\tilde{b}_2h^2)^2 \equiv 0
\end{aligned} \tag{6.62}$$

where $y^P(t_0+h)$ *is the* value calculated in the predictor step and

$$f_1 = f(t_0 + h, y^P(t_0 + h)), \qquad f_0 = f(t_0, y(t_0)),$$
$$f_{-2} = f(t_0 - 2h, y(t_0 - 2h)), \qquad f_{-1} = f(t_0 - h, y(t_0 - h)).$$

After the values for the coefficients $\tilde{a}_0^i$, $\tilde{a}_1^i$, $\tilde{a}_2^i$, $\tilde{b}_1^i$ and $\tilde{b}_2^i$ are obtained, the corrected value of $y^C{}_{i+1}(t)$ is calculated as:

$$y^C{}_{i+1} = \frac{\tilde{a}_0^i + \tilde{a}_1^i h + \tilde{a}_2^i h^2}{1 + \tilde{b}_1^i h + \tilde{b}_2^i h^2} \tag{6.63}$$

This value is then accepted as a good approximation for $y(t)$ and this becomes the initial condition for calculations in the next step.

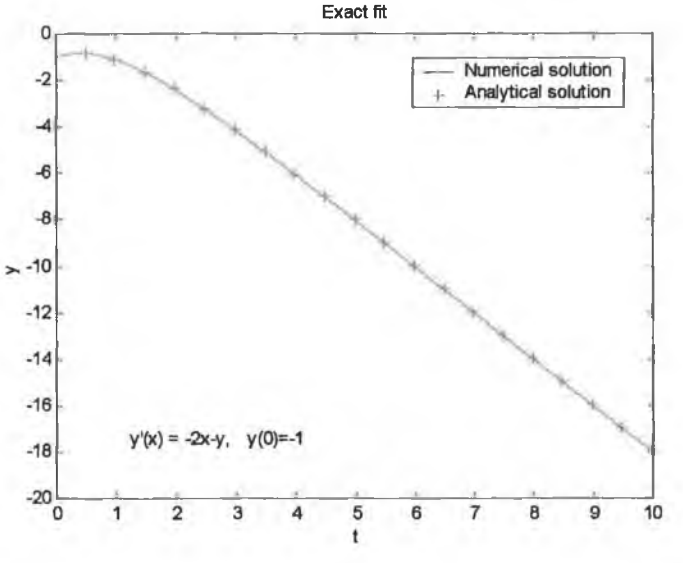


Fig. 6.2. *Exact fit method*

Fig. 6.2 presents the numerical solution for the test problem

$$\frac{dy}{dx} = -2x - y, \quad y(0) = -1 \tag{6.64}$$

compared to its exact analytical solution

$$y(t) = -3e^{-t} - 2x + 2. \tag{6.65}$$

As can be seen, agreement between the analytical and numerical solution to the ordinary differential equation (6.64) is excellent.

The Exact-fit method is a multistep method, which means that it is not self-starting. It requires one of singlestep methods in order to calculate the values for the initial $m+n-1$ time steps, e.g. RKF algorithm, after which calculation is resumed according to the Exact-fit method algorithm.

6.5.2. Padé fit method

The Exact-fit method requires solving a non-linear system of equations to obtain the coefficients a_0, a_1, a_2, b_1 and b_2 at each time-step. This may be computationally expensive when large systems are solved since it involves inverting the Jacobian matrix. In order to avoid solving a non-linear system at the each time step, the Padé-fit method initially fits $y(t)$ with a polynomial of order $m+n$. For the Padé approximant (6.57) this requires a 4th order polynomial:

$$y(t) = C_0 + C_1 t + C_2 t^2 + C_3 t^3 + C_4 t^4 \tag{6.66}$$

and hence,

$$\frac{dy}{dt} = C_1 + 2C_2 t + 3C_3 t^2 + 4C_4 t^3 . \tag{6.67}$$

Taking the current value of $y(t)$ as the initial condition for next step $(t_0 = t_i)$ and fitting dy/dt over the past 4 points $(t_0, t_0\text{-}h, t_0\text{-}2h, t_0\text{-}3h)$ yields a 5x5 system of *linear* equations where C_0, C_1, C_2, C_3 and C_4 are unknowns, i.e.

$$\begin{aligned} C_0 &= y_i(t) \\ C_1 &= f_0 \\ C_1 - 2hC_2 + 3h^2C_3 - 4h^3C_4 &= f_{-1} \\ C_1 - 4hC_2 + 12h^2C_3 - 32h^3C_4 &= f_{-2} \\ C_1 - 6hC_2 + 27h^2C_3 - 108h^3C_4 &= f_{-3} \end{aligned} \tag{6.68}$$

where

$$f_0 = f(t_0, y(t_0)), \qquad f_{-2} = f(t_0 - 2h, y(t_0 - 2h)),$$
$$f_{-1} = f(t_0 - h, y(t_0 - h)), \qquad f_{-3} = f(t_0 - 3h, y(t_0 - 3h)).$$

This system of linear equations can be solved at a fraction of the cost in terms of computational time and resources compared to solving the non-linear system of equations (6.60). Equating the Padé approximant (6.57) with the polynomial approximation (6.66), yields:

$$y(t) = C_0 + C_1 t + C_2 t^2 + C_3 t^3 + C_4 t^4 = \frac{a_0 + a_1 t + a_2 t^2}{1 + b_1 t + b_2 t^2} \tag{6.69}$$

Cross multiplying and collecting the corresponding coefficients up to the 4th order, the following analytical expressions are obtained for the coefficients a_0^i, a_1^i, a_2^i, b_1^i and b_2^i:

$$\begin{aligned} a_0^i &= C_0 \\ a_1^i &= \frac{C_1^2 C_3 - C_1 C_2^2 + C_0 C_2 C_3 - C_0 C_1 C_4}{C_1 C_3 - C_2^2} \\ a_2^i &= \frac{2C_1 C_2 C_3 + C_0 C_2 C_4 - C_1^2 C_4 - C_0 C_3^2 - C_3^3}{C_1 C_3 - C_2^2} \\ b_1^i &= \frac{C_2 C_3 - C_1 C_4}{C_1 C_3 - C_2^2} \\ b_2^i &= \frac{C_2 C_4 - C_3^2}{C_1 C_3 - C_2^2} \end{aligned} \tag{6.70}$$

These coefficients are then used to calculate $y_{i+1}(t)$ as:

$$y_{i+1} = \frac{a_0^i + a_1^i h + a_2^i h^2}{1 + b_1^i h + b_2^i h^2}. \tag{6.71}$$

Implementation of a predictor-corrector algorithm is not complicated with this method either. To find an expression for the corrector dy/dt is fitted over 4 points but this time from $t_0 - 2h$ to $t_0 + h$. Taking the current value of $y(t)$ as the initial condition for the next step $(t_0 = t_i)$ yields another 5x5 system of linear equations where $\tilde{C}_0$, $\tilde{C}_1$, $\tilde{C}_2$, $\tilde{C}_3$ and $\tilde{C}_4$ are the new set of unknown coefficients calculated according to the following formula for a 4th order method:

$$\begin{aligned} \tilde{C}_0 &= y(i) \\ \tilde{C}_1 &= f_0 \\ \tilde{C}_2 &= \frac{1}{12h}(-3f_1 + 8f_0 - 6f_{-1} + f_{-2}) \\ \tilde{C}_3 &= \frac{1}{6h^2}(f_1 - 2f_0 + f_{-1}) \\ \tilde{C}_4 &= \frac{1}{24h^3}(f_1 - 3f_0 + 3f_{-1} - f_{-2}) \end{aligned} \tag{6.72}$$

where $y^P(t_0+h)$ *is the* value calculated in the predictor step and

$$\begin{aligned} f_1 &= f(t_0 + h, y^P(t_0 + h)), & f_0 &= f(t_0, y(t_0)), \\ f_{-2} &= f(t_0 - 2h, y(t_0 - 2h)), & f_{-1} &= f(t_0 - h, y(t_0 - h)). \end{aligned}$$

These coefficients are then used for calculating $\tilde{a}_0$, $\tilde{a}_1$, $\tilde{a}_2$, $\tilde{b}_1$ and $\tilde{b}_2$ in order to obtain $y_{i+1}(t)$, $t = t_0 + h$, in the corrector step, i.e.

$$y^{C}{}_{i+1} = \frac{\tilde{a}_0^i + \tilde{a}_1^i h + \tilde{a}_2^i h^2}{1 + \tilde{b}_1^i h + \tilde{b}_2^i h^2} \tag{6.73}$$

This value is then accepted as a good approximation for *y(t)*, and this becomes the initial condition for calculations in the next step. Fig. 6.3 presents a comparison between the numerical solution calculated using the Padé-fit method for the test problem (6.64) and the exact analytical solution (6.65).

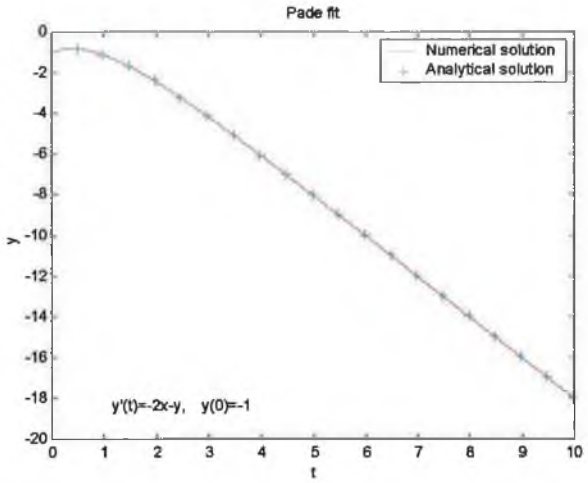
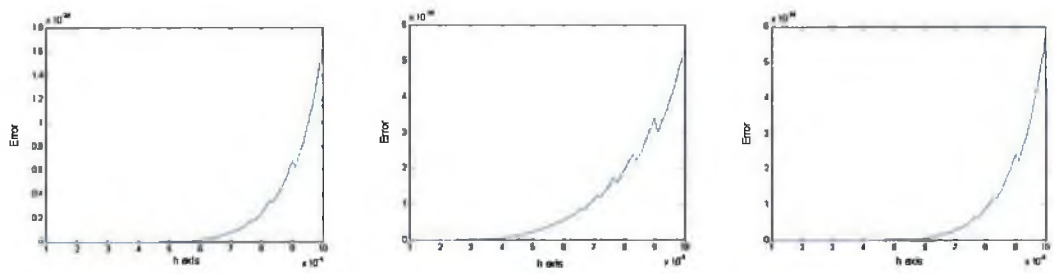


Fig. 6.3. *Padé-fit method*

As with the Exact-fit method, the Padé-fit method is a multistep method and it requires a singlestep method in order to calculate the initial *m+n-1* values after which calculation is resumed according to the Padé fit method algorithm.

6.5.3. Some comments on the Exact and Padé fit methods

Both the Exact-fit and the Padé-fit methods are multistep methods, hence they require one of the singlestep methods in order to calculate starting values. The method used here is the 4th order Runge-Kutta-Felhberg. The Exact-fit method requires solving a system of non-linear equations at each step while the Padé-fit needs only a linear system to be solved. Thus the Padé-fit method is less computationally expensive.



a) Exact-fit method b) Padé-fit method c) Adams-Moulton method

Fig. 6.4. *Error comparison*

Fig. 6.4 shows the mean-square error distribution calculated for 100 different stepsizes h in the range $[10^{-4}, 10^{-3}]$ of the 4th order implementations of the Exact-fit, Padé-fit and Adams-Moulton method for the example problem (6.64). As can be seen, the Exact-fit method is superior in terms of accuracy to the Padé-fit method when compared for different stepsizes h. The Exact-fit method is 9 orders of magnitude more accurate than the Padé-fit method for the same problem (Fig. 6.4.a) and Fig. 6.4.b)) and moreover the Exact-fit method is more than an order of magnitude more accurate than the widely used Adams-Moulton method of the same order (Fig. 6.4. c)). Therefore, the use of the Exact-fit method is suggested when a highly accurate solution is sought, while the Padé-fit method is recommended when computational speed is of the essence.

6.6. Methods that use derivatives of the function $f(t,y(t))$

The Padé-Taylor and the Padé-Xin method are the two **new** methods presented in this section. Both methods require obtaining an analytical expression for derivatives of the function f in order to calculate the coefficients that are necessary for approximation of y. The approximating function for the Padé-Taylor method is again one of the Padé approximants (6.66), but the Padé-Xin method is based on a slightly different approach – the approximating function is a combination of a Padé approximant and exponential part, as given in (6.92).

6.6.1. Padé-Taylor method

The Padé-Taylor method is similar to the Padé-fit method – the difference is in the way the coefficients C_i, $i=0,\ldots,4$ for calculating a_i and b_i in (6.66) are obtained. Assume that (6.4) has unique solution $y(t)$ on $[a,b]$ and that there exist $p+1$ derivatives of $y(t)$ on $[a,b]$. The solution $y(t)$ can be expanded in a Taylor series about any point t_n as:

$$y(t)=y(t_n)+(t-t_n)y'(t_n)+\frac{1}{2!}(t-t_n)^2y''(t_n)+\ldots+\frac{1}{p!}(t-t_n)^p y^{(p)}(t_n)+\frac{(t-t_n)^{p+1}}{(p+1)!}y^{(p+1)}(\xi_n) \quad (6.74)$$

This expansion is valid for $t\in[a,b]$, $t_n<\xi<t$. Substituting $t=t_{n+1}$ and $h=t_{n+1}-t_n$ in (6.74) yields:

$$y(t_{n+1})=y(t_n)+hy'(t_n)+\frac{h^2}{2!}y''(t_n)+\ldots+\frac{h^p}{p!}y^{(p)}(t_n)+\frac{h^{p+1}}{(p+1)!}y^{(p+1)}(\xi_n) \quad (6.75)$$

If equation (6.75) is written in terms of approximate values and taking into account that $y' = f(t, y(t))$, it becomes

$$y_{i+1} = y_i + f(t_i, y_i)h + \frac{f'(t_i, y_i)}{2!}h^2 + \ldots + \frac{f^{(p-1)}(t_i, y_i)}{p!}h^p . \tag{6.76}$$

Now consider a 4th order approximation. Comparing (6.76) to (6.66) one can see that:

$$\begin{aligned} C_0 &= y_i \\ C_1 &= f(t_i, y_i) \\ C_2 &= \frac{f'(t_i, y_i)}{2!} \\ C_3 &= \frac{f''(t_i, y_i)}{3!} \\ C_4 &= \frac{f'''(t_i, y_i)}{4!} \end{aligned} \tag{6.77}$$

Therefore it is possible to substitute the coefficients C_i, $i = 0,\ldots,4$ in (6.70) and proceed with calculations in the manner described for the Padé fit method. The advantage of the Padé-Taylor method is that it is not necessary to solve any system of equations in order to obtain the coefficients C_i, $i = 0,\ldots,4$. However, it is necessary to be able to obtain derivatives of high order in analytical form.

The implementation of a predictor-corrector algorithm is different to the previously described methods. It is necessary to develop a corrector step, which further increases the accuracy of the method but avoids the necessity for calculation of even higher order derivatives. The proposed corrector for the 4th order method is as follows:

$$y_{i+1}^C = y_{i+1}^P - \frac{h^5}{720} f'''(t_{i+1}, y_{i+1}^P) \tag{6.78}$$

where y_{i+1}^P is obtained from the predictor stage and is an estimate of $y(t_{i+1})$, the true solution at time t_{i+1}. The rationale for the choice of corrector is as follows: Consider the simple Forward Euler, which is an explicit method, given as:

$$y_{n+1} = y_n + hf(t_n, y_n) \tag{6.79}$$

On the other hand, the Trapezoidal method, which is implicit in nature, is given:

$$y_{n+1} = y_n + \frac{h}{2}(f(t_{n+1}, y_{n+1}) + f(t_n, y_n)) . \tag{6.80}$$

Now consider a predictor-corrector method that uses the Forward Euler as a predictor and uses the Trapezoidal method as a corrector. If this predictor-corrector scheme is applied to the test function $y = e^{-t}$, it is observed that the result is equivalent to initially employing an explicit second-order Taylor series expansion, i.e.

$$y_{i+1} = y_i + hf(t_i, y_i) + \frac{h^2}{2!} f'(t_i, y_i) \tag{6.81}$$

Returning to the method specified in equations (6.79) and (6.80), a Padé approximate of order n matches the first n+1 coefficients (time-domain moments) of a Taylor series expansion. It also provides additional terms. Considering the test function e^{-t}, a fourth order Padé approximate is given by:

$$y_{i+1} = \phi_{2/2} = \frac{1 - 0.5h + 0.0833h^2}{1 + 0.5h + 0.0833h^2} \tag{6.82}$$

This function matches the first five coefficients of a Taylor series expansion for e^{-t}. It also produces additional terms, the first of which is:

$$T_6 = \frac{-1}{144} h^5 \tag{6.83}$$

However, the correct sixth coefficient in a Taylor series expansion for e^{-t} is:

$$T_6 = \frac{-1}{120} h^5 \tag{6.84}$$

Noting the observation regarding the Euler predictor-corrector, a corrector is chosen so as to match T_6 for the particular test function, $y = e^{-t}$, without requiring a higher-order derivative. Hence, the choice of corrector specified in equation (6.78).

As an illustrative example, the following well-known classic equation system is used [S97]. Equations (6.85) constitute a stiff system of differential equations.

$$\begin{aligned} \frac{du}{dt} &= 998u + 1998v, \quad \text{u(0)=1.0} \\ \frac{dv}{dt} &= -999u - 1999v, \quad \text{v(0)=1.0} \end{aligned} \tag{6.85}$$

The analytical solution for system in equation (6.85) is given by:

$$\begin{aligned} u(t) &= 4e^{-t} - 3e^{-1000t} \\ v(t) &= -2e^{-t} + 3e^{-1000t} \end{aligned} \tag{6.86}$$

The result computed with the Adams Moulton predictor-corrector for a step-size of 1ms, superimposed on the analytical solution, is shown in Fig. 6.5. Note that there is a discrepancy between the Adams-Moulton method result and the exact result over the time interval from 4-6 seconds. The corresponding result computed with the new predictor-corrector and a step-size of 8ms, i.e. **eight** times larger, is shown in Fig. 6.6. Note the increased level of accuracy.

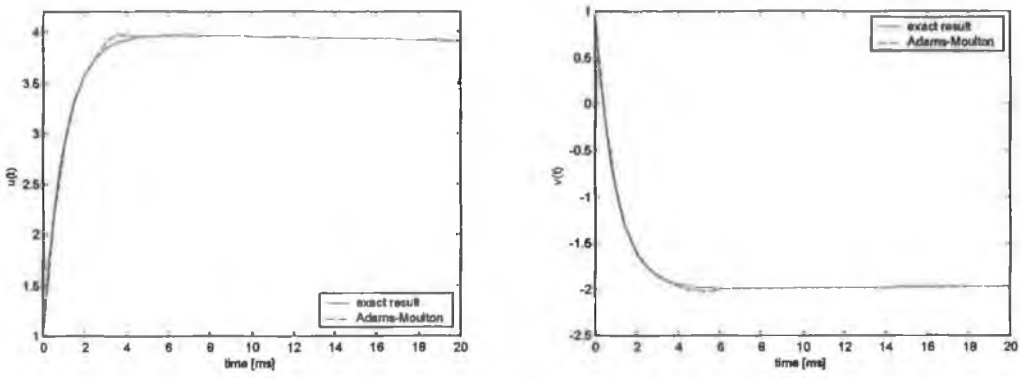


Fig. 6.5. *Results computed with Adams Moulton predictor corrector*

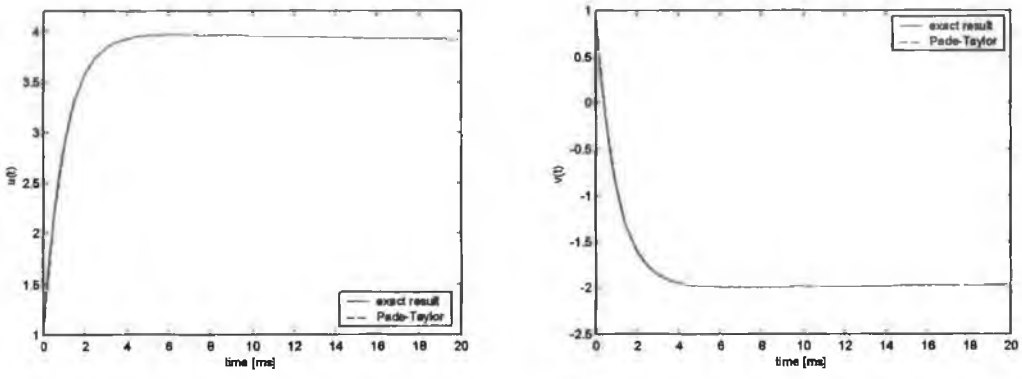


Fig.6.6. *Results computed with Padé-Taylor predictor corrector*

As evidenced by these results, the new technique is superior for the given step-size and therefore permits a significantly larger step-size for a comparable level of accuracy. A similar speed up (seven times), is obtained for a MESFET amplifier circuit described by a system of ten stiff differential equations as reported in [CDB02].

6.6.2. Padé-Xin method

This method also uses derivatives of the function $f(t, y(t))$ in (6.55) to calculate the required coefficients in order to obtain a satisfactory approximation. It combines the Padé approximation and the approach introduced by Wu [W98]. The basic idea is that the theoretical solution to a stiff system can be represented locally in the interval $[t_i, t_{i+1}]$ by the composition of a polynomial and exponential function

$$PE(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_k t^k + b_1 e^{b_2 t} = P_k(t) + b_1 e^{b_2 t}, \qquad (6.87)$$

where a_i, $i = 0, 1, \ldots, k$ and b_j, $j = 1, 2$ are coefficients to be determined and

$$P_k(t) = a_0 + a_1 t + a_2 t^2 + \ldots + a_k t^k = \sum_{l=0}^{k} a_l t^k$$

is the polynomial of k^{th} order.

First, assume that the approximation y_i of the solution $y(t)$ is written as the composition of a polynomial and an exponential function (6.87), i.e.

$$y_i = PE(t_i) = P_k(t) + b_1 e^{b_2 t} \tag{6.88}$$

and that the function f is $k+1$ times differentiable on $[t_i, t_{i+1}]$. Using (6.88) and the following $(k+2)$ conditions:

$$f_i^{(n)} = PE^{(n)}(t_i), \quad n = 0,1,...,k+1 \tag{6.89}$$

yields a system of $(k+3)$ algebraic equations from which it is possible to calculate the required coefficients. Since Padé approximants have better approximating properties than polynomials of the same order [GW99], the next step is to equate the polynomial part with the Padé approximant, i.e.

$$P_k(t) = \frac{p_0 + p_1 t + ... p_{k/2} t^{k/2}}{1 + q_1 t + ... q_{k/2} t^{k/2}} \tag{6.90}$$

p_l, q_l, $l=0,1,..,k/2$ are the Padé coefficients and q_0 is set to be one ($q_0=1$). For the purpose of keeping the notation simple, it is assumed that k is an even number although for this method it is not a necessary condition. This procedure then yields a new expression for y_i

$$y_i = PX(t_i), \tag{6.91}$$

where

$$PX(t) = \frac{p_0 + p_1 t + ... p_{k/2} t^{k/2}}{1 + q_1 t + ... q_{k/2} t^{k/2}} + b_1 e^{b_2 t}. \tag{6.92}$$

The exponential part in (6.92) is used to extract the behaviour of the fast changing part of the solution. A low-order Padé approximant is all that is necessary for the slower changing part.

Therefore, by setting $k = 2$, equation (6.87) becomes:

$$PE(t) = a_0 + a_1 t + a_2 t^2 + b_1 e^{b_2 t} \tag{6.93}$$

Using (6.87) and (6.91) the following system of algebraic equations are obtained:

$$\begin{aligned}
&a_0 + a_1 t_i + a_2 t_i^2 + b_1 e^{b_2 t_i} = y_i \\
&a_1 + 2a_2 t_i + b_1 b_2 e^{b_2 t_i} = f_i \\
&2a_2 + b_1 b_2^2 e^{b_2 t_i} = f_i' \\
&b_1 b_2^3 e^{b_2 t_i} = f_i'' \\
&b_1 b_2^4 e^{b_2 t_i} = f_i'''
\end{aligned} \tag{6.94}$$

from which one can calculate the required coefficients as:

$$
\begin{aligned}
b_2 &= \frac{f_n'''}{f_n''} = Z_n \\
b_1 &= \frac{f_n''}{Z_n^{\,3}} e^{-b_2 t_n} \\
a_2 &= \frac{1}{2}(f_n' - \frac{f_n''}{Z_n}) \\
a_1 &= (f_n - \frac{f_n''}{Z_n^{\,2}}) - (f_n' - \frac{f_n''}{Z_n}) t_n \\
a_0 &= (y_n - \frac{f_n''}{Z_n^{\,3}}) - (f_n - \frac{f_n''}{Z_n^{\,2}}) t_n + \frac{1}{2}(f_n' - \frac{f_n''}{Z_n}) t_n^{\,2}
\end{aligned}
\tag{6.95}
$$

The equations in (6.95) give *direct* analytical formula for calculating the coefficients of Padé approximant, without the need to employ Newton algorithm or Gaussian elimination. The only operations involved are basic numerical operations, i.e. addition and multiplication.

Equating the Padé approximant to the polynomial part $P_k(t)$ and matching coefficients up to k^{th} order (in this example $k = 2$)

$$
a_0 + a_1 t + a_2 t^2 = \frac{p_0 + p_1 t}{1 + q_1 t} \tag{6.96}
$$

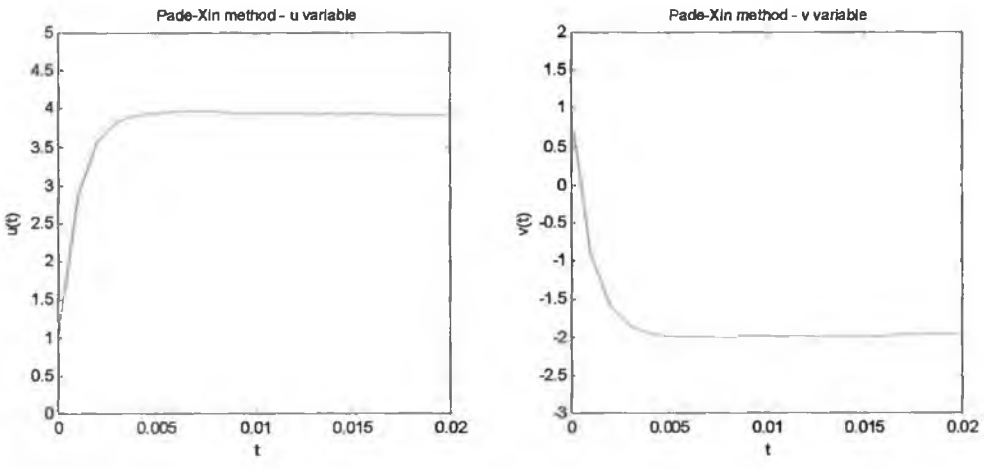
yields following relationship for the Padé approximant coefficients:

$$
\begin{aligned}
p_0 &= a_0 \\
p_1 &= \frac{a_1^{\,2} - a_0 a_2}{a_1}. \\
q_1 &= -\frac{a_2}{a_1}
\end{aligned}
\tag{6.97}
$$

Finally, the value for y_{i+1} is obtained from:

$$
y_{i+1} = PX(t_{i+1} - t_i) = PX(h) = \frac{p_0 + p_1 h}{1 + q_1 h} + b_1 e^{b_2 h} \tag{6.98}
$$

Fig. 6.7 presents results calculated for the system of stiff differential equations in (6.85) using the Padé-Xin method.



a) Padé-Xin method – u variable *b) Padé-Xin method – v variable*

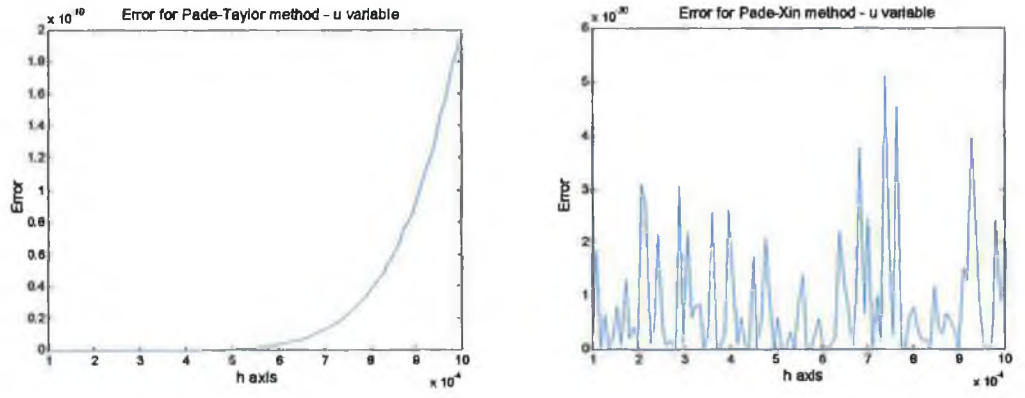
Fig.6.7. *Results computed with Padé-Xin method*

As with the Padé-Taylor method, the Padé-Xin method also requires finding analytical expressions for high-order derivatives of the function $f(t,y(t))$. This enables utilisation of direct analytical formulae (6.95) and (6.97) to calculate approximation coefficients without the need to solve a system of (non)linear equations as is case with the Exact-fit and Padé-fit method. Hence, a considerable speed-up in simulation may be achieved.

6.6.3. Some comments on Padé-Taylor and Padé-Xin methods

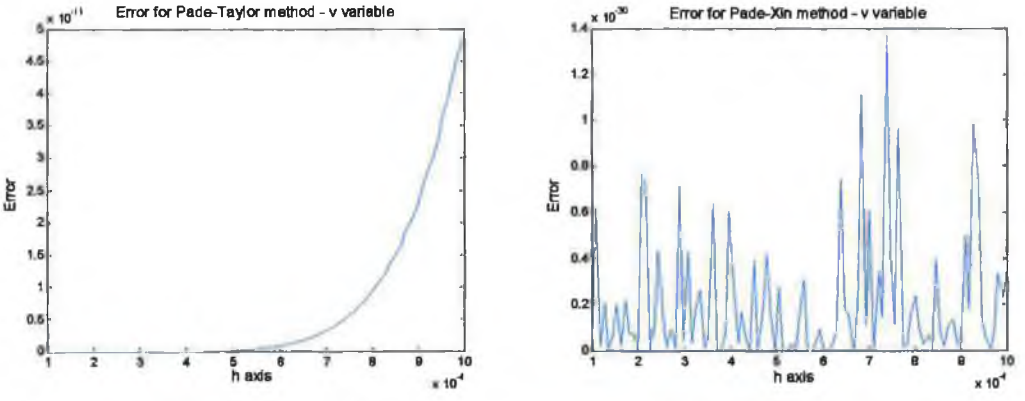
Unlike the Exact-fit or Padé-fit methods, the Padé-Taylor and Padé-Xin methods require finding analytically n^{th}-order derivatives of an n-times differentiable function $f(t,y(t))$ in (6.55). However, this is not a serious problem as it is always possible to obtain derivatives in analytical form for any such function. This is done in order to eliminate the computationally expensive process of solving the (non)linear system of equations associated with obtaining the Padé coefficients in the Exact-fit or Padé-fit methods. Instead, *analytical* formulae for the coefficients in the Padé approximation are readily available and the computational cost associated with their evaluation is negligible. Thus both the Padé-Taylor and Padé-Xin methods are found to be highly computationally efficient.

Fig. 6.8 shows the mean-square error distribution for 10 different stepsizes h in range $[10^{-4}, 10^{-3}]$ for variable u. Fig. 6.9. shows the same for variable v.



a) Pade-Taylor method error *b) Pade-Xin method error*

Fig. 6.8. *Mean-square error for variable u*



a) Pade-Taylor method error *b) Pade-Xin method error*

Fig. 6.9. *Mean-square error for variable v*

As can be seen the error in the Padé-Xin method of the order 10^{-30} and it achieves excellent accuracy when compared to all other methods.

6.7. A comparison between the presented numerical methods and conclusions

In total, **four new** methods for solution of the stiff IVP have been presented in this Chapter. The basic idea behind these methods is similar to that of [GN97], where a sequence of local approximations to $y(t)$ is built in order to provide a solution to the IVP as defined in (6.55). These local approximations are Padé approximates as given in (6.56). The method is then advanced in time by using the solution at time t as the initial

condition for the next time-step, i.e. for calculating y_{i+1} the solution for y_i is the initial condition.

The Exact fit and Padé fit methods are multistep methods. Therefore, they require use of a singlestep algorithm, e.g. RK family of methods, to calculate the several first values. These methods do not require finding derivatives of the function $f\left(t, y(t)\right)$ in (6.4) thus making these methods attractive to use when it is complicated or impossible to find the corresponding derivatives. The development of a corrector formula is straightforward for these two methods.

The Padé-Taylor and Padé-Xin are singlestep methods, i.e. self-starting methods. They require obtaining derivatives of the function $f\left(t, y(t)\right)$ in analytical form, but in turn, this enables derivation of formulae for the coefficients of Padé approximant. Hence, these methods are an accurate alternative when derivatives are readily available. In addition, corrector formula for use with the Padé-Taylor method is also developed and presented here.

The Padé-Xin method proved to be the most accurate out of all four methods proposed. The methods were tested both on a single ODE and a small system of ODEs. Implementation of the proposed methods for large-scale systems is the next step in algorithm development. After this, all algorithms may be compared time-wise and the most suitable one in terms of efficiency and accuracy will be chosen for implementation in a circuit simulator.

CHAPTER 7

Wavelets in Relation to Envelope Transient Simulation

In the reminder of this dissertation, a novel wavelet-based envelope transient analysis technique for the simulation of highly non-linear circuits subjected to high-frequency modulated signals is presented. The underlying principle behind wavelets and the Wavelet Theory (WT) is described in this Chapter. The relation between Wavelet Theory and the classical Fourier Transform (FT) and its derivative, the Short Term Fourier Transform (STFT), is illustrated. The main properties of the WT are summarised from several great resources [M93], [B98], [D92], [V99], [P04], [M98], [BGG98], [K94], [CW96]. Finally, a detailed description of the wavelet-like basis especially suited for dealing with strong non-linearities is given. This wavelet basis will then be used in the novel wavelet-based envelope transient analysis method presented in Chapter 8.

7.1. Introduction

Wavelet theory (WT) is a relatively recent area of scientific research that emerged during the 20th century from the study of Calderon-Zygmund operators in mathematics, the study of the theory of subband coding in engineering and the study of renormalisation group theory in physics. Ingrid Daubechies's paper entitled *"Orthonormal bases of compactly supported wavelets"* [D88] provided the starting point for development of modern wavelet theory. She suggested the use of the *orthonormal* bases of the form $2^{j/2}\psi(2^j x-k), j,k \in \mathbb{Z}$, where $\psi(x)$ is *continuous* and has a *continuous derivative* and *compact support.* This led to an avalanche of interest in wavelets, both from a theoretical and practical viewpoint. Today her *Ten lectures on wavelets* [D92] are the theoretical basis for any researcher entering the wavelet world. More recent work by Donoho [D93], Coifman [CW92] and many others have given theoretical explanations as to why wavelets work so well on such a broad range of problems in image and signal processing, mathematical modelling, numerical analysis and electrical engineering.

The application area of wavelets is enormous and versatile [M93], [WT04], [SSW02], [Y92], [P93], [BGG98], [CW96]. In image and signal processing, wavelets are used for applications such as signal and image compression, nonlinear filtering (denoising), statistical estimation, calculation of FFTs and for the approximation of a function in terms of scaling functions and wavelets in a computationally efficient manner. In numerical analysis, wavelets have proven themselves as a useful tool for solving operator equations (i.e. matrix equations, the differential form of Maxwell's equations, the multi-scale moment method for solving integral equation), obtaining numerical solution to boundary value problems and for solving large real and sparse or complex matrix equations. In electrical engineering, wavelets are used for electromagnetic scattering and radiation problems, packaging and interconnects, linear system modelling and non-linear semiconductor device modelling. Wavelets are also used in real-world applications, the compression of the FBI fingerprints data and in medical and biomedical signal and image processing, e.g. in micro-potential extraction in ECG, noise removal in ECG or for identifying a quick transitory signal in EEGs. Even very specialised areas like fractals, turbulence theory, oceanography, seismic and geophysical signal processing, astronomy, metallurgy, finance and even internet traffic description have found wavelets a very promising and exciting new tool. The list of wavelet applications is by no means exhausted; new research results in different scientific areas are published every day.

7.2. The rationale for wavelets

To explain the rationale for wavelets, first consider the very simple example of a general representation of a real number x expressed in the form:

$$x = \sum_{i=-\infty}^{+\infty} d_i b^i , \tag{7.1}$$

where b is the base. For example, for decimal system representation $b=10$, for binary $b=2$ and for hexadecimal $b=16$. So, if the representation system is known in advance, it is only necessary to know the digits d_i to fully represent the number. Theoretically, the number of digits can be finite (e.g. for rational real numbers) or infinite (e.g. for irrational real numbers). However, for all practical calculation purposes, only a finite number of digits is considered and only the accuracy requirement dictates how many digits are actually taken.

In the function domain, all classical expansions of the Laurent, Taylor or Fourier type reflect this approach exactly. For example, the Fourier Expansion states that any 2π-periodic function $f(x)$ can be written as:

$$f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) , \tag{7.2}$$

i.e., the sum of sine and cosine functions multiplied by certain coefficients $(k \geq 1)$:

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(x)dx, \quad a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx. \tag{7.3}$$

Therefore, for Fourier theory, the sine and cosine functions represent the basis in which any 2π-periodic function $f(x)$ may be represented. It is only necessary to identify coefficients a_i, $i=0,1,...$ and b_j, $j=1,2,...$ to have complete information about $f(x)$.

In Wavelet Theory, instead of sine and cosine functions, a different set of basis functions with some predetermined properties is constructed and used in function representation. So the wavelet basis $\Psi = \{\psi_I : I \in \mathcal{J}\}$, where $\mathcal{J}$ is suitable chosen index set, is in a sense just another collection of basis functions that may be used to expand a given function $f(x)$ as:

$$f = \sum_{I \in \mathcal{J}} d_I(f)\psi_I \tag{7.4}$$

Therefore, such an expansion associates with a function f, the array $\boldsymbol{d} = \{d_I(f)\}_{I \in \mathcal{J}}$ of coefficients, as was the case for the classical expansions. However, there are a few points by which a wavelet expansion differs from the classical expansions of Taylor or Fourier [D99]:

1. A Taylor expansion places strong demands on the regularity of f such as analyticity, while (7.4) is typically valid for a much larger class of functions such as square integrable ones, i.e. the only requirement is that the series on the right hand side of (7.4) converges in the corresponding norm.
2. More importantly, the digits d_I convey *very detailed information* on f due to the structure of $\mathcal{J}$. Each $\mathcal{J}$ comprises information of a different type such as scale and spatial location. For example, classical wavelets on the real line denoted by t are generated by scaling (determined by value of j) and translating (determined by value of k) a single function ψ, i.e. $\psi_{j,k} = 2^{j/2}\psi(2^j t - k)$. Therefore, by

noting $I \leftrightarrow (j,k)$, it is possible to associate ψ_I with detail information determined by j around a spatial location encoded by k.

3. In contrast to approximating the solution function f of a given operator equation on some mesh (of fixed highest resolution), wavelet based schemes aim to determine its representation with respect to a basis. This means that during the solution process, wavelet based algorithms will track only those coefficients in the unknown array ***d(f)*** in (7.4) that are the *most significant* for approximating f with as few as possible degrees of freedom. This property contributes immensely towards the efficiency of such algorithms.

These essential differences explain the use of wavelet-based expansions in various applications. The main aim of most today's wavelet researchers is to create a set of *expansion functions* and *transforms* that give a useful and efficient description of a function or signal. A set of expansion functions does not have to be a *basis* for that function space but, for most applications, it is a desirable property since a set of basis functions has some nice qualities in terms of efficiency of calculations and elegance in analytical representation.

7.3. From Fourier Transform (FT) to Wavelet Transform (WT)

The development of wavelet theory has been largely influenced by Fourier techniques and harmonic analysis concepts as presented in Section 7.2. Fourier techniques provide a simple construction of a function (7.2) and also an elegant means of performing analysis in the frequency domain [IJ02], [OSB99]. For many years, it has been an invaluable tool for many signal-processing engineers who needed to obtain the frequency content of an analysed signal. However, although by far a superior technique when dealing with frequency domain applications, the Fourier Transform is practically useless if time-domain information is required, as is the case for mixed linear/non-linear applications.

7.3.1. Fourier Transform (FT)

The Fourier Transform (FT) is a very well known and a widely used signal-processing technique [IJ02], [OSB99]. The FT decomposes a signal in terms of complex exponential functions of different frequencies. The Fourier Transform pair is given by:

$$X(f) = \int_{-\infty}^{+\infty} x(t)\, e^{-2j\pi f t} dt \tag{7.5}$$

$$x(t) = \int_{-\infty}^{+\infty} X(f)\, e^{2j\pi f t} df \tag{7.6}$$

where x denotes a signal in the time domain, and X denotes the signal in the frequency domain. Equation (7.5) is called the *Fourier transform of* $x(t)$ and (7.6) is called the *inverse Fourier transform of* $X(f)$. The above transform pair has many useful properties that can be found in [IJ02], [OSB99] (or in any of the vast array of resources dealing with the theory of mathematical expansions or signal processing. They will not be given here).

Upon closer inspection of (7.5), it can be seen that the signal $x(t)$ is multiplied by an exponential term at some frequency f and then integrated over *all times*. If the result of this integration is a large value, the signal $x(t)$ has a dominant spectral component at frequency f. Conversely, if the integration result is a small value, the contribution of the signal component of frequency f is negligible, and if the integration result is zero, than the signal does not have the component with frequency f at all. If the integration process is repeated for all frequencies $f \in (-\infty, +\infty)$, the *frequency spectrum* of signal $x(t)$ is obtained. In essence, the frequency spectrum of a signal shows which frequencies are present in the signal and the amplitude of the component with frequency f determines the amount that component contributes to the signal.

The important thing to note about the FT is the fact that the integration in (7.5) is from minus infinity to plus infinity in the time domain. It follows that whether the frequency component f appears at time t_1 or t_2 in the signal, it will have the same effect on the integration. Therefore, the Fourier Transform gives all the necessary information about the frequency spectrum of a signal but cannot give any information about its time-domain properties. This is the big disadvantage of the Fourier Transformation. The FT has only frequency resolution and no time resolution, i.e. it is possible to determine all the frequencies present in the signal but it is not possible to determine *when* they are present [P04]. This is not a drawback if the analyzed signal is a *stationary signal*, i.e. the signal whose frequency content does not change in time. Therefore, when analyzing a stationary signal, one is only interested in the frequency content since all frequency components present in the signal exist at all times.

However, there are many very important *non-stationary signals*, e.g. ECG (electrocardiograph, electrical activity of the heart) or EEG (electroencephalograph,

electrical activity of the brain) to name just a few. In fact, for most naturally occurring signals, the frequency content does change in time, hence they are non-stationary signals. For non-stationary signals, the frequency components do not appear at all times and it may be of great importance to have information about their occurrence in time, something that the FT just is not capable of providing.

7.3.2. The Short Term Fourier Transform (STFT)

In the past few decades, several techniques that enable signal representation in both the time and frequency domain at the same time have been developed. The core idea behind these techniques is to cut the signal into a number of parts and then to analyse each part separately. This will, obviously, give more information about when/where the frequency component appears. But, this also introduces a new problem, i.e. the choice of the technique to be used to cut the signal.

7.3.2.1. The Dirac pulse as a window

One approach is to use a very short time window using a Dirac pulse of form

$$f(t-t_0)=\begin{cases}1, & t=t_0\\ 0, & t\neq t_0\end{cases}, \tag{7.7}$$

which "will" give all the frequency components present at a moment $t=t_0$. However, this is not true and the explanation of why this is not possible lies in the fact that cutting the signal corresponds to a convolution between the signal and the cutting window. Since convolution in the time domain is identical to multiplication in the frequency domain and since the Fourier transform of a Dirac pulse contains all possible frequencies, the frequency components of the signal will be smeared out over the entire frequency axis [P04]. That means that contrary to the standard Fourier transform result, excellent time resolution has been achieved but all frequency resolution has been lost! This situation is analogous to the Heisenberg uncertainty principle in physics, which states that it is not possible to know both the exact position and exact momentum of a particle in any moment of time. The better information about the particle's position at a time point t_0 is available, the less accurate is information about its momentum at t_0. In terms of signal processing, this principle can be reformulated as: It is not possible to exactly know *what frequency exists* at *what time instance*, rather it is possible only to know what **frequency bands** exist at what **time intervals** [P04].

7.3.2.2. The Short Term Fourier Transform (STFT)

The Short Term Fourier Transform (STFT) was developed as an answer to the problem of cutting up a signal with the aim to obtain time localization of the signal. In the STFT, the signal is divided into portions deemed small enough, so that the signal may be assumed stationary. For this purpose, a window function w is chosen in a manner such that the width of this window must be equal to the segment where the analysed non-stationary signal may be assumed to be stationary.

The STFT is defined by:

$$STFT_X^{\mathrm{w}}(\tau, f) = \int_t [x(t)\mathrm{w}(t-\tau)]e^{-j2\pi f t}dt\,. \tag{7.8}$$

As can be seen from (7.8), the STFT is in essence the FT of a multiplication between the signal $x(t)$ and window function $w(t)$. Assuming that the window length, L, is chosen such that it can be assumed that the signal is stationary on it, then the obtained result will be a true frequency representation of the signal that is also localised in time. Localisation in time is determined by the parameter τ, which specifies the mid point of a window function.

However, by taking the STFT of the signal, the individual exact frequency components that exist in the signal are not identified anymore. Rather a 'band' of frequencies present is determined. The narrower the window length is, the better the time resolution that is obtained but the frequency resolution is worse. If a wider window length is chosen, the time resolution is poorer but the frequency resolution is improved. Finally, if an infinite window length is chosen, the STFT is reduced to the FT with perfect frequency resolution, but no time resolution whatsoever.

So with the STFT, the problem of cutting up the analysed signal is essentially the problem of choosing the window length parameter [P04]. A narrow window gives good time resolution, but the frequency resolution is poor. On the other hand, wide windows will give good frequency resolution but the time localisation is worse. Furthermore, wide windows may also violate the condition of stationarity of the signal. If the signal at hand has well separated frequency components, than the frequency resolution may be sacrificed and a narrow window may be chosen in order to obtain good time localisation. But if the signal is a complex one, with various frequency components, choosing a constant length window can produce some misleading results. This is the biggest disadvantage of the STFT. Once the window length, L, also known as *the support of the window*, is chosen, it has to remain constant for all analysis. This

inflexibility of resolution in both the time and frequency domain is overcome by introducing the new type of transformation, namely the Wavelet Transform (WT).

7.3.3. The Wavelet Transform (WT)

The Wavelet Transform is the transform developed to overcome the problem of the fixed resolution of the STFT. As with the STFT, the WT is capable of providing simultaneously the time and frequency information of a signal, hence giving a *time-frequency representation* of the signal. However, using a *fully scalable modulated window,* the WT gives *variable resolution* at all times, while the STFT can only give fixed resolution determined by the chosen support of the window used [P04]. This window is translated along the signal and for every position the spectrum is calculated. Then the size of window is changed (scaled) and the spectrum is once again calculated. At the end, a collection of time-frequency representations of the signal is obtained all with different resolutions. Thus, wavelet analysis is often referred to as *Multiresolution Analysis (MRA).* When discussing the Wavelet Transformation the term *"time-scale representation"* is used, where scale is in a way the opposite of frequency. This is due to the fact that the term 'frequency' is usually reserved for the Fourier Transform.

When analyzing a signal whose frequency spectrum is composed of both low and high frequencies (e.g. envelope modulated signal), less relative error will be made if the high frequency components can be located in time and the low frequency component is better resolved in the frequency domain [P04]. Therefore, the STFT is inappropriate for analyzing such signals since it resolves every spectral component with equal resolution and does not take into account if the component is at the high or low end of the frequency spectrum. On the other hand, the Wavelet Transform is capable of analyzing the spectrum of the signal with different resolution. This is due to the concept of *Multiresolution Analysis (MRA),* which is designed to *give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies* [P04]. ***This is an excellent approach for the envelope-modulated signal that has a low frequency component for a long duration and a high frequency component for a short period of time.***

7.4. Wavelets and Wavelet Transform (WT)

The name *wavelet* or *ondelette* was coined in the early 1980's by French researchers [MAF+82], [GM84] meaning the *small wave. Smallness* refers to the fact

that this window function is compactly supported (i.e. the window is of finite length) and *wave* refers to the condition that it has to be an oscillatory function. The important properties of wavelets and conditions are briefly discussed in the rest of this Section. A short description of both the Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) is given [V99], [P04].

7.4.1. The Continuous Wavelet Transform (CWT)

The Continuous Wavelet Transform (CWT), given in (7.9), is a function of two variables, the translation parameter τ and the scaling parameter s.

$$\Psi_x^{\psi}(\tau,s)=\frac{1}{\sqrt{s}}\int x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt \tag{7.9}$$

$$x(t)=\frac{1}{C_{\psi}^2}\int_s\int_{\tau}\Psi_x^{\psi}(\tau,s)\frac{1}{s^2}\psi\left(\frac{t-\tau}{s}\right)d\tau ds\,. \tag{7.10}$$

The function $\psi_{s,\tau}(t)$, given as:

$$\psi_{s,\tau}(t)=\frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{7.11}$$

is the transforming function and is analogous to the window function w in STFT. For completeness, equation (7.10) gives the Inverse Continuous Wavelet Transform used for reconstruction of the original signal $x(t)$ under the condition that the wavelet transform has been taken with the respect to the same mother wavelet [Y93].

As can be seen, the CWT is performed in a similar manner to the STFT, i.e. the signal is multiplied with the function ψ (this time called wavelet rather than window function) and the CWT is then computed separately for different segments of the time-domain signal. However, the major difference compared to the STFT and the crucial property of the CWT is that *the width of window is different* for every single spectral component. It is this property that enables the WT to provide different resolutions at all times.

The *mother wavelet* $\psi(t)$ is in essence the *prototype* [P04] for generating all window functions. All windows employed are either dilated or compressed (by factor s) and shifted (parameter τ) versions of this mother wavelet. The factor $1/\sqrt{s}$ is used for energy normalization across different scales. The *translation parameter* τ is related to the location of the window which is shifted through the signal. Its role in the WT is similar to that in the STFT.

The *scaling parameter s*, or the *scale*, is related to frequency in that it is the frequency inverse. Therefore, the high scale corresponds to low frequencies or a global view of the signal and low scales correspond to high frequencies or a detailed view of the signal. Therefore, the STFT is characterized by translation and frequency parameters, and the WT with translation and scale parameters.

7.4.2. Wavelet Properties

The two most important wavelet properties are *admissibility*, which stipulates that the wavelet function $\psi(t)$ must be a wave and the *regularity conditions* that state that the wavelet transform should decrease quickly with decreasing scale parameter *s*.

7.4.2.1. Admissibility condition

It can be shown [S96], [V99] that square integrable functions $\psi(t)$ satisfying the admissibility condition:

$$\int \frac{\left|\mathbb{F}_{\psi}(\omega)\right|^2}{|\omega|} d\omega < +\infty, \tag{7.12}$$

can be used to first decompose (analysis) and then reconstruct (synthesis) a signal without loss of information. Here, $\mathbb{F}_{\psi}(\omega)$ stands for the Fourier transform of $\psi(t)$.

There are two important notes regarding the admissibility condition (7.12) :

1. Wavelet $\psi(t)$ must have a *band-pass spectrum* since (7.12) can be fulfilled only if the Fourier transform $\mathbb{F}_{\psi}(\omega)$ vanishes at zero frequency, i.e.

$$\left|\mathbb{F}_{\psi}(\omega)\right|^2_{\omega=0} = 0. \tag{7.13}$$

2. Wavelet $\psi(t)$ must be a *wave* since the zero at the zero frequency in the frequency domain also means that in the time domain the average value of the wavelet must be zero, i.e.

$$\int \psi(t) dt = 0, \tag{7.14}$$

and therefore, it must be oscillatory function.

7.4.2.2. Regularity conditions

Regularity of a function is a quite complex concept and more about it can be found in [D92], [BGG98]. Basically, the regularity conditions state that wavelet function $\psi(t)$ should have some smoothness and concentration in both the time and

frequency domains. Valens [V99], used the *vanishing moments* concept to relay the basic idea behind regularity conditions. The p^{th} moment M_p of a wavelet can be defined as:

$$M_p = \int t^p \psi(t) dt . \tag{7.15}$$

A Taylor series expansion for the continuous wavelet transform $\Psi_x^{\psi}(\tau,s)$ in (7.9) at $t=0$ until order n with $\tau=0$ (for simplicity) is [S96]:

$$\Psi_x^{\psi}(0,s) = \frac{1}{\sqrt{s}}\left[\sum_{p=0}^{n} x^p(0) \int \frac{t^p}{p!} \psi\left(\frac{t}{s}\right) dt + O(n+1)\right]. \tag{7.16}$$

If (7.16) is now rewritten in terms of moments, the following approximation is obtained:

$$\Psi_x^{\psi}(0,s) = \frac{1}{\sqrt{s}}\left[x(0)M_0 s + \frac{x^{(1)}(0)}{1!} M_1 s^2 + \ldots + \frac{x^{(n)}(0)}{n!} M_n s^{n+1} + O(s^{n+2})\right]. \tag{7.17}$$

For wavelets fulfilling the admissibility condition, the 0^{th} moment is equal to zero since

$$M_0 = \int \psi(t) dt = 0 . \tag{7.18}$$

If the wavelet is chosen such as $M_i = 0,\ i = 1,\ldots,n$, then from (7.17) it follows that the wavelet transform coefficients $\Psi_x^{\psi}(\tau,s)$ will decay as fast as s^{n+2} for a smooth signal $x(t)$. Therefore, if the wavelet has n vanishing moments, then the approximation order of the wavelet transform is also n. For practical purposes, the required number of vanishing moments is strongly influenced by the application at hand. Also, the moments do not have to be exactly zero, very often a small value is good enough.

7.4.3. The Discrete Wavelet Transform (DWT)

Although the CWT is a very important part of Wavelet Theory, it is very difficult to implement (7.9) or (7.10) in practice due to three key properties of the CWT [V99]:

- The CWT is highly *redundant* since its calculations are based on a set of continuously scalable functions which do not form an orthogonal basis. For most practical purposes redundancy is not desirable, although sometimes it may be put to a good use, e.g. denoising applications.
- There are an *infinite* number of wavelets in the CWT and for any practical calculation, the number has to be finite.
- For most functions, the CWT has *no analytical solution* and they have to be calculated numerically. This requires very fast algorithms to exploit the advantages of the wavelet transforms.

7.4.3.1. Redundancy

To address the problem of redundancy, it is desirable to form an orthogonal basis of wavelets. Daubechies [D92] introduced the modified wavelet representation of (7.11) which is called the *discrete wavelet*:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j} \right). \tag{7.19}$$

A discrete wavelet is normally a (piecewise) continuous function and discreteness refers to the fact that discrete wavelets are not continuously scalable and translatable functions but can only be scaled and translate in discrete steps determined by integers j and k. The translation factor τ_0 depends on a fixed dilation step $s_0 > 1$. Usually s_0 is chosen as $s_0 = 2$ so that the sampling of the frequency axis correspondents to dyadic sampling, and τ_0 is usually chosen as $\tau_0 = 1$ to enable dyadic sampling of the time axis also. In that case, (7.19) can be written as:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi\left(2^{-j} t - k\right). \tag{7.20}$$

Using a discrete wavelet to transform a continuous signal, a series of wavelet coefficients analogous to Fourier coefficients (7.3) are obtained and the decomposition is referred to as the *wavelet series decomposition*. Daubechies [D92] has proven that the necessary and sufficient condition for stable reconstruction from wavelet series decomposition is that the energy of the wavelet coefficients must lie between two positive bounds, i.e.

$$A\|f\|^2 \le \sum_{j,k} \left| \langle f, \psi_{j,k} \rangle \right|^2 \le B\|f\|^2 \tag{7.21}$$

where $\|f\|^2$ is the energy of $f(t)$, $A > 0$, $B < \infty$ and A, B are independent of $f(t)$. When (7.21) is satisfied, the family of basis functions $\psi_{j,k}$ with $j, k \in \mathbb{Z}$ is referred to as a *frame* with *frame bounds* A and B. If $A = B$, the frame is called a *tight frame* and the discrete wavelets behave exactly like an orthonormal basis. If $A \neq B$, the frame is called a *dual frame* and reconstruction is still possible although the decomposition wavelet is then different from the reconstruction wavelet.

If the mother wavelet is chosen such that its discrete wavelets are orthogonal to their own dilations and translations, i.e.

$$\int \psi_{j,k}(t)\ \psi^*_{m,n}(t)\ \mathrm{dt} = \begin{cases} 1, j = m \text{ and } k = n \\ 0, \text{otherwise} \end{cases} \tag{7.22}$$

then an arbitrary signal can be reconstructed by summing the orthogonal basis functions, weighted by the wavelet transform coefficients [S96] :

$$x(t) = \sum_{j,k} \Psi_x^{\psi}(j,k)\psi_{j,k}(t) . \tag{7.23}$$

Equation (7.23) is the *inverse wavelet transform for discrete wavelets* and it shows how the original signal can be reconstructed.

Therefore, if the orthogonal discrete wavelets are used, the redundancy is removed enabling efficient calculations. However, the wavelets need not be orthogonal in order to represent a signal. Sometimes, redundancy can be very helpful, e.g. to reduce sensitivity to noise [S96].

7.4.3.2. Finite number of wavelets

The signal $x(t)$ has to have a finite energy to enable its frequency spectrum and time duration to be covered with wavelets. Mathematically, this condition can be stated as:

$$\int |x(t)|^2 \, dt < \infty \tag{7.24}$$

that is, the L^2-norm of the signal $x(t)$ should be finite. For natural signals, this condition is always fulfilled since they normally have finite energy.

As regards the translation of the wavelets, this is limited by the duration of the signal $x(t)$. As regards the dilation parameter, there is the problem of an infinite number of scales needed to cover the entire signals spectrum. From the admissibility condition (7.13) it is known that the wavelets have a band pass spectrum. So wavelets should be designed in such a manner that they touch each other at the end of their spectra if the signal's spectrum is to be entirely covered [V99], as shown in Fig. 7.1.

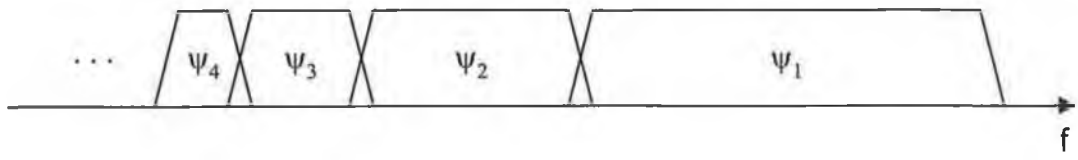


Fig. 7.1. *Touching wavelet spectra resulting from scaling of the mother wavelet in the time domain*

However, the introduction of every new scale means that half of the remainder of spectrum is covered. Thus, the second half has to be covered. In other words, the spectrum will never be entirely covered because it is necessary to cover an infinite number of remainder half-intervals (this is analogous to the famous paradox of Zeno).

The solution that was proposed by Mallat [M88] is in form of low-pass filter that covers the remainder of spectrum all the way to zero, as shown in Fig. 7.2.

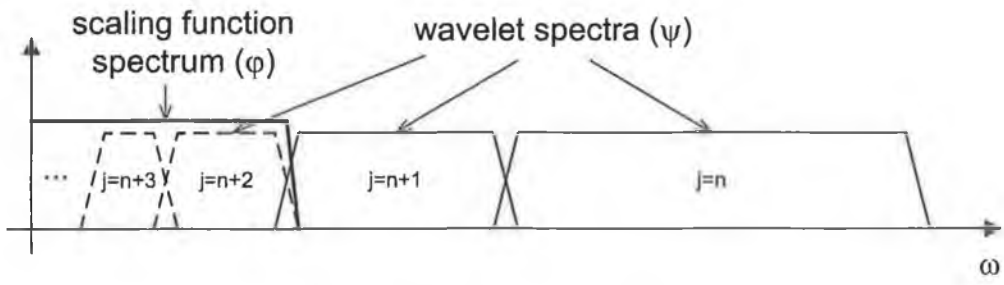


Fig. 7.2. *Scaling and wavelet function spectra*

This low-pass filter corresponds to the so-called *scaling function* $\varphi(t)$ that, like any other signal, can be represented by a wavelet decomposition up to a scale j as:

$$\varphi(t) = \sum_{j,k} \Psi_{\varphi}^{\psi}(j,k)\psi_{j,k}(t) \tag{7.25}$$

Therefore, using a combination of scaling functions and wavelets, the number of wavelets is in effect reduced from an infinite number to a finite number. From a signal representation point of view, there is no loss of information but from a wavelet analysis point of view, information about the scales up to level $j = n+1$ is lost. For scaling functions, it is possible to state a sort of admissibility condition similar to (7.12) as:

$$\int \varphi(t)dt = 1, \tag{7.26}$$

which shows that the 0^{th} moment of scaling function cannot vanish.

From Fig. 7.2, it is clear that the scaling function can be considered as a low-pass filter and the set of wavelets as a high-pass filter bank. This leads to a fast algorithm for computing the CWT.

7.4.3.3. Fast algorithm for CWT

A fast algorithm for calculating the wavelet transform is based on the fact that the Wavelet Transform may be considered as a filter bank. Then, the Wavelet Transform is analogous to passing the signal through this filter bank, a technique analogous to the idea of *subband coding* in signal processing theory. The outputs of the different filter stages are the wavelet and scaling function transform coefficients.

Fig. 7.3 presents the idea of an iterated filter bank where a signal is first passed through a low-pass (scaling) and a high-pass filter (wavelets). Although the filter corresponding to the wavelets is called high-pass in reality, it is a band-pass filter due to

the limited bandwidth of natural signals. The wavelet coefficients at this stage will identify the highest level of detail and the rest of the information will be contained in the output of low pass filter. If more details are required then this output can be passed again through low- and high-pass filters to produce another set of wavelet coefficients corresponding to the next level of detail and the rest of the information is again contained in the output of the low pass filter. Each time the signal is passed through a pair of filters, the frequency spectrum of the output is halved as can be seen from Fig. 7.3. These iterations may be continued until the required results are obtained.

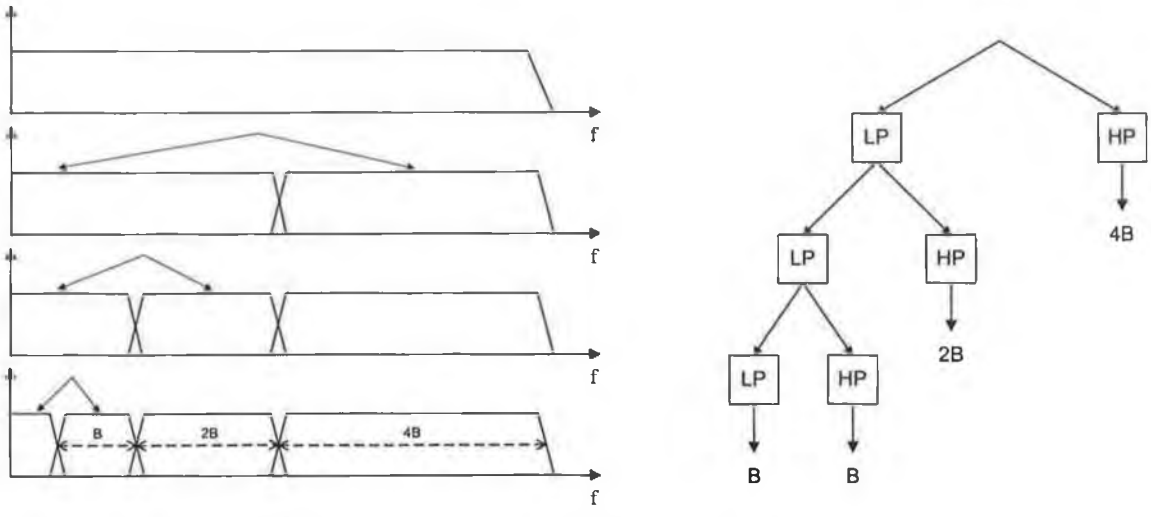


Fig. 7.3. *Splitting the signal spectrum with an iterated filter bank*

Mallat [M88] was the first to discover this analogy between a wavelet transform and a subband coding scheme. Since the output of different filtering stages are the scaling and wavelet coefficients at different resolutions this kind of signal analysis is often referred to as a *multiresolution analysis (MRA).*

Developing the DWT has solved the practical implementation issues that emerged from the CWT. The undesirable redundancy of the CWT has been removed via the introduction of orthogonal discrete wavelets, and the introduction of scaling functions has enabled the reduction of an infinite number of wavelets to a finite number. The remarkable property of the DWT is that the implementation can be done without specifying any wavelet, which means that the problem of the non-existence of analytical solutions of the CWT is not an issue anymore.

7.5. A wavelet-like multiresolution collocation technique

As was seen in Section 7.4 the critical property of wavelet theory is that it presents a general framework for wavelets and wavelet transforms. Within this framework it is possible to design an actual wavelet system with properties best suited for a particular problem at hand. In this section, a wavelet-like system for solving the particular category of partial differential equations (PDEs) that comprise initial boundary value problems, as proposed by Cai and Wang [CW96], will be presented. The key points and properties of this wavelet system are briefly summarised. This particular wavelet-like system has been chosen because of its superior capabilities in dealing with strong non-linearities [CW96]. This wavelet system is then used to develop a highly efficient wavelet-based envelope transient simulation technique as will be presented in the Chapter 8.

7.5.1. Introduction

Wavelet based methods for solving partial differential equations (PDE) have recently caught the attention of researchers developing efficient numerical techniques. [CW96], [DK097], [ML03], [C03], [HKG+03], [GLR+90], [XS92]. Because of their properties of good localisation in both the time and frequency domain [BGG98], [V99], [P04], [S96], as well as the ability to choose at which scale of detail the wavelet approximation will be used, the wavelet based multiresolution scheme can be seen as a potentially excellent approach to efficiently obtain solutions that vary dramatically in both the time and frequency domain. For example, in communication circuits when a high-frequency RF carrier is modulated by a low-frequency information signal, the circuit waveforms vary significantly both in the time and frequency domain. Thus a wavelet based solution scheme is a natural approach since *a high-frequency RF carrier may be resolved by wavelet functions while scaling functions may be used to efficiently capture the behaviour of the low-frequency information signal.*

Most of the wavelet-based schemes for PDEs use Daubechies' orthonormal wavelets for wavelet decomposition of $L^2(\mathbb{R})$ [LT90], [LRT91], [WA94], [AW95]. However, the strong non-linearities as encountered in most high-frequency systems can require many wavelet levels for an adequate representation and thus seriously degrade the inherent advantages of the wavelet schemes.

Cai and Wang propose the direct construction of a multiresolution analysis (MRA) scheme for a Sobolev space $H_0^2(I)$ defined on a bounded interval I. Such a MRA $(V_0 \subset V_1 \subset V_2 \cdots)$ enables the decomposition of $H_0^2(I)$ into the form:

$$H_0^2(I) = V_0 \oplus_{j=0}^{\infty} W_j, \tag{7.27}$$

where $\oplus$ stands for the orthogonal direct sum. W_j denotes the orthogonal compliment of V_j in the space V_{j+1} where:

$$V_{j+1} = V_j \oplus W_j. \tag{7.28}$$

V_j can be generated from the scaling functions, defined later in (7.29) and (7.30), through dilations and translations. Also most of the basis for all of the subspaces W_j can be generated from the "mother wavelet" function through dilations and translations of (7.36). However, the two boundary functions in each W_j are generated from another function located at the boundary ("boundary mother wavelet" defined later in (7.37)) by dilations and reflections. Since the inner product considered in this case is in the space $H_0^2(I)$ and not $L^2(I)$, both of these mother wavelet functions will no longer have vanishing moments of the first two orders as is usual for wavelets. Therefore, strictly speaking, the scaling and wavelet function used here do not represent a usual wavelet system. However, the projection f_j of any function $f \in H_0^2(I)$ on V_j still provides a "general picture" of the function f while the projection on W_j keeps its local details [CW96]. Hence, the magnitude of the coefficients in the wavelet expansion of functions in $H_0^2(I)$ does reflect the local scales and changes of the function to be approximated [CW96]. Bearing this in mind, from this point onwards, the functions that form the bases for W_j will be referred to as "wavelets" but with the understanding that they differ from ordinary wavelets with their non-vanishing moments.

7.5.2. Scaling functions $\varphi(x)$ and $\varphi_b(x)$

As mentioned before, the MRA for the Sobolev space $H_0^2(I)$ is generated using two types of functions: interior and boundary functions.

The *interior scaling function* is given by :

$$\varphi(x) = N_4(x) = \frac{1}{6}\sum_{j=0}^{4}\binom{4}{j}(-1)^j (x-j)_+^3 \tag{7.29}$$

and the *boundary scaling function* is:

$$\varphi_b(x)=\frac{3}{2}x_+^2-\frac{11}{12}x_+^3+\frac{3}{2}(x-1)_+^3-\frac{3}{4}(x-2)_+^3+\frac{1}{6}(x-3)_+^3 \tag{7.30}$$

where $N_4(x)$ is a fourth-order B-spline and for any real number n, x_+^n is defined as:

$$x_+^n=\begin{cases}x^n, & t\geq 0\\ 0, & otherwise\end{cases}. \tag{7.31}$$

As a pair, (7.29) and (7.30) satisfy the *two-scale relationship*:

$$\begin{aligned}\varphi(x)&=\sum_{k=0}^{4}2^{-3}\binom{4}{k}\varphi(2x-k)\\ \varphi_b(x)&=\beta_{-1}\varphi_b(2x)+\sum_{k=0}^{2}\beta_k\varphi(2x-k)\end{aligned}, \tag{7.32}$$

where $\beta_{-1}=\frac{1}{4},\beta_0=\frac{11}{16},\beta_1=\frac{1}{2},\beta_2=\frac{1}{8}$. Some important properties of $\varphi(x)$ and $\varphi_b(x)$ may be summarised as:

$$\begin{aligned}&i)\quad supp(\varphi(x))=[0,4];\\ &ii)\quad supp(\varphi_b(x))=[0,3];\\ &iii)\quad \varphi(x),\varphi_b(x)\in H_0^2(I);\\ &iv)\quad \varphi'(1)=-\varphi'(3)=\frac{1}{2},\varphi'(2)=0,\varphi_b'(1)=\frac{1}{4},\varphi_b'(2)=-\frac{1}{2};\\ &v)\quad \varphi(1)=\varphi(3)=\frac{1}{6},\varphi(2)=\frac{2}{3},\varphi_b(1)=\frac{7}{12},\varphi_b(2)=\frac{1}{6}.\end{aligned} \tag{7.33}$$

Now it is possible to define V_j, for any $j,k\in\mathbb{Z}$, as the linear span of:

$$V_j=span\left\{\varphi_{j,k}(x)\,|\,0\leq k\leq 2^jL-4;\varphi_{b,j}(x),\varphi_{b,j}(L-x)\right\}, \tag{7.34}$$

where

$$\begin{aligned}\varphi_{j,k}(x)&=\varphi(2^jx-k)\\ \varphi_{b,j}(x)&=\varphi(2^jx)\end{aligned}. \tag{7.35}$$

The $V_j, j\in\mathbb{Z}^+$ form a *MultiResolution Analysis* (MRA) for $H_0^2(I)$ equipped with the norm given in Appendix F in the following sense [CW96]:

1) $V_0\subset V_1\subset V_2\cdots$;
2) $clos_{H_0^2}(\bigcup_{j\in\mathbb{Z}^+}V_j)=H_0^2(I)$;
3) $\bigcap_{j\in\mathbb{Z}^+}V_j=V_0$
4) for each j, $\left\{\varphi_{j,k}(x),\varphi_{b',j}(x),\varphi_{b,j}(L-x)\right\}$ is a basis of V_j.

7.5.3. Wavelet functions $\psi(x)$ and $\psi_b(x)$

Two interior and boundary wavelet functions are used. The *interior mother wavelet function* is given by:

$$\psi(x)=-\frac{3}{7}\varphi(2x)+\frac{12}{7}\varphi(2x-1)-\frac{3}{7}\varphi(2x-2) \tag{7.36}$$

and the *boundary mother wavelet function* is:

$$\psi_b(x)=\frac{24}{13}\varphi_b(2x)-\frac{6}{13}\varphi(2x) \tag{7.37}$$

Some important properties of $\psi(x)$ and $\psi_b(x)$ are given as:

i) $\psi(x),\psi_b(x)\in V_1$;

ii) $\psi(n)=\psi_b(n)=0,\ \forall n\in\mathbb{Z}$.

The following *wavelets* are defined as:

$$\psi_{j,k}(x)=\psi(2^j x-k),\quad j\geq 0,k=0,\ldots,n_j-3 \tag{7.38}$$

$$\psi^l_{b,j}(x)=\psi_b(2^j x),\quad \psi^r_{b,j}(x)=\psi_b(2^j(L-x)), \tag{7.39}$$

where $n_j=2^j L$. To simplify notation, let

$$\psi^l_{b,j}(x)=\psi_{j,-1}(x),\quad \psi^r_{b,j}(x)=\psi_{j,n_j-2}(x). \tag{7.40}$$

Therefore, when $k=-1$ or $k=n_j-2$, the wavelet functions $\psi_{j,k}(x)$ actually denote the two boundary wavelet functions.

Now, for each $j\geq 0$ it is possible to define W_j as:

$$W_j=span\left\{\psi_{j,k}(x)\,|\,k=-1,\ldots,n_j-2\right\}, \tag{7.41}$$

which is the orthogonal compliment of V_j in V_{j+1} under the inner product defined in Appendix F. This may be denoted as

$$V_{j+1}=V_j\oplus W_j,\quad j\in\mathbb{Z}^+ \tag{7.42}$$

where $\oplus$ stands for

a) $V_j\perp W_j$ under the inner product;

b) $V_{j+1}=V_j+W_j$.

Cai and Wang showed that the following two properties hold [CW96]:

1) $W_j\perp W_{j+1}, j\in\mathbb{Z}^+$;

2) $H_0^2(I)=V_0\oplus_{j\in\mathbb{Z}^+}W_j$.

Therefore, any function $f(x)\in H_0^2(I)$ may be approximated by a function $f_j(x)\in V_j$

at any chosen level $j \geq 0$. The function $f_j(x) \in V_j = V_0 \oplus W_0 \oplus W_1 \oplus \cdots W_{j-1}$ has a unique orthogonal decomposition

$$f_j(x) = f_0 + g_0 + g_1 + \ldots + g_{j-1}, \tag{7.43}$$

where $f_0 \in V_0$, $g_i \in W_i$, $0 \leq i \leq j\text{-}1$.

7.5.4. Spline functions $\eta(x)$

To deal with non-homogenities at boundaries, Cai and Wang [CW96] introduced the interpolating spline $\boldsymbol{I}_{b,j} f(x), j \geq 0$

$$\boldsymbol{I}_{b,j} f(x) = \alpha_1 \eta_1(2^j x) + \alpha_2 \eta_2(2^j x) + \alpha_3 \eta_2(2^j (L-x)) + \alpha_4 \eta_1(2^j (L-x)) \tag{7.44}$$

where

$$\begin{aligned} \eta_1(t) &= (1-t)_+^3 \\ \eta_2(t) &= 2t_+ - 3t_+^2 + \frac{7}{6} t_+^3 - \frac{4}{3}(t-1)_+^3 + \frac{1}{6}(t-2)_+^3 . \end{aligned} \tag{7.45}$$

The coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are determined by, so called, *end conditions*. There are two common types of the end conditions: derivative end and not-a-knot conditions.

The *derivative end conditions* are:

$$\begin{aligned} \boldsymbol{I}_{b,j} f(0) &= f(0), & \boldsymbol{I}_{b,j} f(L) &= f(L), \\ (\boldsymbol{I}_{b,j} f)'(0) &= f'(0), & (\boldsymbol{I}_{b,j} f)'(L) &= f'(L). \end{aligned} \tag{7.46}$$

The coefficients $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ obtained from these conditions yield

$$f(x) - \boldsymbol{I}_{b,j} f(x) \in H_0^2(I)$$

and thus the decomposition can be applied, i.e.

$$f_j(x) = \boldsymbol{I}_{b,j} f(x) + f_0 + g_0 + g_1 + \ldots + g_{j-1}, \tag{7.47}$$

where $f_0 \in V_0$, $g_i \in W_i$, $0 \leq i \leq j\text{-}1$. Therefore, any function $f(x) \in H^2(I)$ may be approximated by a function $f_j(x)$. The order of the approximation is $O(2^{-4j})$ and it depends on the chosen j [CW96].

The *not-a-knot conditions* [CW96], [DB78] given in (7.48) impose the restriction that a spline $\boldsymbol{I}_{b,j} f(x)$ agrees with the function $f(x)$ at one additional point (τ_1 and τ_2) near each boundary.

$$\begin{aligned} \boldsymbol{I}_{b,j} f(0) &= f(0), & \boldsymbol{I}_{b,j} f(L) &= f(L), \\ \boldsymbol{I}_{b,j} f(\tau_1) &= f(\tau_1), & \boldsymbol{I}_{b,j} f(\tau_2) &= f(\tau_2) \end{aligned} \tag{7.48}$$

The reasoning behind this is that if the solution varies dramatically near a boundary, as

is often case, interpolating with a spline $\boldsymbol{I}_{b,j}f(x)$ whose coefficients $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ are defined through *derivative* end conditions (7.46) will yield unacceptably large errors. The interpolating spline $f_j(x)$ defined as in (7.47) through *not-a-knot* conditions still has an approximation to $f(x)$ of the same order $O(2^{-4j})$ [CW96], [SV72] but $f(x)-\boldsymbol{I}_{b,j}f(x)$ is no longer in the space $H_0^2(I)$.

7.5.5. Interpolant operators I_{V_0} and I_{W_j} and the wavelet interpolation $\mathcal{P}_J f(x)$

The *cubic interpolant* $\boldsymbol{I}_{V_0}f(x)$ in V_0 for a function $f(x)\in H_0^2(I)$ is defined as:

$$\boldsymbol{I}_{V_0}f(x)=c_{-1}\varphi_b(x)+\sum_{k=0}^{L-4}c_k\varphi_{0,k}(x)+c_{L-3}\varphi_b(L-x), \tag{7.49}$$

where the following statement holds

$$\boldsymbol{I}_{V_0}f(x_k^{(-1)})=f_k^{(-1)},\quad k=1,\ldots,L-1. \tag{7.50}$$

$f_k^{(-1)}=f(x_k^{(-1)}),\ k=1,\ldots,L-1$ are the values of function $f(x)$ at the interior collocation points $x_k^{(-1)}$ in V_0 defined as:

$$x_k^{(-1)}=k,\quad k=1,\ldots,L-1. \tag{7.51}$$

If (7.49) and (7.50) are combined and then written for all collocation points (7.51), the following matrix equation can be written:

$$\boldsymbol{f}^{(-1)}=\boldsymbol{B}\boldsymbol{c}, \tag{7.52}$$

where

$$\boldsymbol{f}^{(-1)}=\begin{bmatrix}f_1^{(-1)}\\ \vdots\\ f_{L-1}^{(-1)}\end{bmatrix}\quad\text{and}\quad \boldsymbol{c}=\begin{bmatrix}c_{-1}\\ \vdots\\ c_{L-3}\end{bmatrix}.$$

Matrix **B** is a constant transform matrix between the values $\boldsymbol{f}^{(-1)}$ and the coefficients $\boldsymbol{c}$. Bearing in mind property (7.33), it may be written as:

$$\boldsymbol{B}=\begin{bmatrix}7/12 & 1/6 & & & & & & \\ 1/6 & 2/3 & 1/6 & & & & & \\ & 1/6 & 2/3 & 1/6 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & 1/6 & 2/3 & 1/6 & & \\ & & & & 1/6 & 2/3 & 1/6 & \\ & & & & & 1/6 & 7/12 & \end{bmatrix}. \tag{7.53}$$

To calculate coefficients $c_k, -1 \le k \le L-3$, the tridiagonal system (7.52) may be solved at the expense of $(8L)$ calculations [CW96].

The *interpolant* $\boldsymbol{I}_{W_j} f(x)$ in W_j, $j \ge 0$ for function $f(x) \in H_0^2(I)$ is defined as:

$$\boldsymbol{I}_{W_j} f(x) = \sum_{k=-1}^{n_j-2} \hat{f}_{j,k} \psi_{j,k}(x), \tag{7.54}$$

where the following statement holds:

$$\boldsymbol{I}_{W_j} f(x_k^{(j)}) = f_k^{(j)}, \quad -1 \le k \le n_j - 2. \tag{7.55}$$

$n_j = 2^j L$ is the dimension of space W_j. $f_k^{(j)} = f(x_k^{(j)}), j \ge 0, -1 \le k \le n_j - 2$ are the values of function $f(x)$ at interior collocation points $x_k^{(j)}$ in W_j defined as:

$$x_k^{(j)} = \frac{k + \frac{3}{2}}{2^j}, \quad j \ge 0, -1 \le k \le n_j - 2. \tag{7.56}$$

Here the not-a-knot conditions are used with $\tau_1\big|_{k=-1} = \frac{1}{2^{j+1}}$, and $\tau_2\big|_{k=n_j-2} = L - \frac{1}{2^{j+1}}$.

If (7.54) and (7.55) are combined and then written for all collocation points (7.56), the following matrix equation can be written

$$\boldsymbol{f}^{(j)} = \boldsymbol{M}_j \hat{\boldsymbol{f}}^{(j)}, \tag{7.57}$$

where

$$\boldsymbol{f}^{(j)} = \begin{bmatrix} f(x_{-1}^{(j)}) \\ \vdots \\ f(x_{n_j-2}^{(j)}) \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{f}}^{(j)} = \begin{bmatrix} \hat{f}_{j,-1} \\ \vdots \\ \hat{f}_{j,n_j-2} \end{bmatrix}.$$

Matrix $\boldsymbol{M}_j$ is a constant transform matrix between values $\boldsymbol{f}^{(j)}$ and coefficients $\hat{\boldsymbol{f}}_{j,k}$, and it may be written [CW96] as:

$$\boldsymbol{M}_j = \begin{bmatrix} 1 & -1/14 & & & & & & \\ -1/13 & 1 & -1/14 & & & & & \\ & -1/14 & 1 & -1/14 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & -1/14 & 1 & -1/14 & & \\ & & & & -1/14 & 1 & -1/13 & \\ & & & & & -1/14 & 1 & \end{bmatrix}. \tag{7.58}$$

To calculate the coefficients $\hat{f}_{j,k}, -1 \le k \le n_j - 2$, again the tridiagonal system (7.57) may be solved at the expense of $(8n_j)$ calculations [CW96].

Assume that the function $f(x) \in H_0^2(I)$ is defined on all collocation points given in (7.51) and (7.56). For all $J\text{-}1 \geq 0$, the *Wavelet interpolation* $\mathcal{P}_J f(x)$ ($\mathcal{P}_J f(x) \in V_0 \oplus W_0 \oplus W_1 \oplus \cdots \oplus W_{J-1}$) is defined as:

$$\mathcal{P}_J f(x) = \hat{f}_{-1,-1}\varphi_b(x) + \sum_{k=0}^{L-4} \hat{f}_{-1,k}\varphi_{0,k}(x) + \hat{f}_{-1,L-3}\varphi_b(L-x) + \sum_{j=0}^{J-1}\left[\sum_{k=-1}^{n_j-2} \hat{f}_{j,k}\psi_{j,k}(x)\right] \quad (7.59)$$

Introducing the following notation:

$$f_{-1}(x) = \boldsymbol{I}_{V_0} f(x) \in V_0 \quad \text{and} \quad f_j(x) = \boldsymbol{I}_{W_j} f(x) = \sum_{k=-1}^{n_j-2} \hat{f}_{j,k}\psi_{j,k}(x) \in Wj, j \geq 0 \quad (7.60)$$

(7.59) can be written as:

$$\mathcal{P}_J f(x) = f_{-1}(x) + \sum_{j=0}^{J-1} f_j(x). \quad (7.61)$$

Cai and Wang [CW96] showed that the following interpolating conditions hold:

$$\begin{aligned} &\mathcal{P}_J f(x_k^{(-1)}) = f(x_k^{(-1)}), \quad 1 \leq k \leq L\text{-}1, \\ &\mathcal{P}_J f(x_k^{(j)}) = f(x_k^{(j)}), \quad j \geq 0, \ -1 \leq k \leq n_j\text{-}2. \end{aligned} \quad (7.62)$$

Equation (7.61) gives the form of the interpolant that may be used to approximate an unknown quantity defined on a certain interval $[0, L]$, $L > 4$. L is taken to be greater than 4 so as to include at least one non-boundary scaling function in (7.59).

7.5.6. Discrete Wavelet Transform (DWT)

Finally, in this section, an algorithm for calculation of the Discrete Wavelet Transform for the chosen wavelet-like system is given. For a compact presentation, the values of $f(x)$ on all interpolation points:

$$\begin{aligned} \boldsymbol{f}^{(-1)} &= \left[f(x_1^{(-1)}), f(x_2^{(-1)}), \ldots, f(x_{L-1}^{(-1)})\right], \\ \boldsymbol{f}^{(j)} &= \left[f(x_{-1}^{(j)}), f(x_0^{(j)}), \ldots, f(x_{n_j-2}^{(j)})\right], \quad j \geq 0. \end{aligned} \quad (7.63)$$

are used to form a vector $\boldsymbol{f} = \left[\boldsymbol{f}^{(-1)}, \boldsymbol{f}^{(0)}, \ldots, \boldsymbol{f}^{(J-1)}\right]^T$. The wavelet coefficients in the expansion (7.59) are given as:

$$\begin{aligned} \hat{\boldsymbol{f}}^{(-1)} &= \left[\hat{f}_{-1,1}, \hat{f}_{-1,2}, \ldots, \hat{f}_{-1,L-1}\right], \\ \hat{\boldsymbol{f}}^{(j)} &= \left[\hat{f}_{j,-1}, \hat{f}_{j,0}, \ldots, \hat{f}_{j,n_j-2}\right], \quad j \geq 0. \end{aligned} \quad (7.64)$$

and are collected into a single vector $\hat{\boldsymbol{f}} = \left[\hat{\boldsymbol{f}}^{(-1)}, \hat{\boldsymbol{f}}^{(0)}, \ldots, \hat{\boldsymbol{f}}^{(J-1)}\right]^T$. The total number of collocation points is denoted by N and given as:

$$N = (L-1) + \sum_{j=0}^{J-1} n_j = 2^J L - 1. \tag{7.65}$$

Depending on the application at hand it may be required to obtain the set of coefficients $\hat{f}$ from the known function f. This process is called the *direct DWT*. On the other hand, the *inverse DWT* involves obtaining values for the function f based on the known values of the coefficients $\hat{f}$ that are usually the output of some processing scheme.

DWT for $f \to \hat{f}$

This direction of transformation will be termed the *direct DWT*. It involves decomposing the function $\boldsymbol{f}$ into a set of coefficients $\hat{\boldsymbol{f}}$ on different resolution levels determined by the parameter J. Since this can involve a great number of coefficients, the aim is to use a fast DWT algorithm to reduce the computing time. In order to obtain a fast DWT, Cai and Wang [CW96] used the following *point value vanishing property* of the functions $\psi_{j,k}(x)$, i.e. for $j > i, -1 \le k \le n_j - 2$:

$$\begin{aligned} &\psi_{j,k}(x_k^{(j)}) = 1, \\ &\psi_{j,k}(x_l^{(i)}) = 0, \quad 1 \le l \le L-1, \quad \text{if } i = -1; \\ &\qquad\qquad -1 \le l \le n_i - 2, \text{ if } i \ge 0; \end{aligned} \tag{7.66}$$

In order to form a wavelet decomposition $\mathcal{P}_J f(x)$ as in (7.61), it is necessary to define expressions for $f_{-1}(x)$ and $f_j(x), 0 \le j \le J-1$.

First, define

$$f_{-1}(x) = \boldsymbol{I}_{V_0} \boldsymbol{f}^{(-1)} = \hat{f}_{-1,-1}\varphi_b(x) + \sum_{k=0}^{L-4} \hat{f}_{-1,k}\varphi_{0,k}(x) + \hat{f}_{-1,L-3}\varphi_b(L-x) \tag{7.67}$$

as an expansion of function $f(x)$ based purely on scaling functions φ. To enable compact notation, the scaling level will be noted as $j = -1$. From (7.50), it follows that $f_{-1}(x)$ interpolates $f(x)$ at the interpolation points $x_k^{(-1)}$, i.e.

$$f_{-1}(x_k^{(-1)}) = f(x_k^{(-1)}), \quad -1 \le k \le L-1 \tag{7.68}$$

and the function $f(x)$ is simply approximated as:

$$f(x) \approx f_{-1}(x) = \boldsymbol{I}_{V_0} f(x). \tag{7.69}$$

Secondly, stepping up one level of detail ($j = 0$), define $f_0(x)$ as:

$$f_0(x) = \boldsymbol{I}_{W_0}(\boldsymbol{f}^{(0)} - (\boldsymbol{I}_{V_0} f)^{(0)}) = \sum_{l=-1}^{n_0-2} \hat{f}_{0,l}\psi_{0,l}(x). \tag{7.70}$$

In essence, $f_0(x)$ is a wavelet decomposition of the *difference* between the complete function presentation $\boldsymbol{f}^{(0)}$ at level $j=0$ and the approximation of $f(x)$ on a previous level ($j=-1$). In this instance, that is simply the cubic interpolant $(\boldsymbol{I}_{V_0} f)^{(0)} = \left\{\boldsymbol{I}_{V_0} f(x_k^{(0)}), -1 \le k \le n_0 - 2\right\}$ evaluated at collocation points $\left\{x_k^{(0)}, -1 \le k \le n_0 - 2\right\}$.

A very important property of functions $f_{-1}(x)$ and $f_0(x)$ defined in (7.67) and (7.70) is that the function $f_{-1}(x)+f_0(x)$ *interpolates f(x) on both* $\left\{x_k^{(-1)}, 1 \le k \le L-1\right\}$ *and* $\left\{x_k^{(0)}, -1 \le k \le n_j - 2\right\}$. This is due to the point value vanishing property (7.66) of the wavelet function $\psi_{0,l}(x)$, i.e.

$$\psi_{0,l}(x_k^{(-1)}) = 0, \quad -1 \le l \le n_0 - 2,\ 1 \le k \le L-1, \tag{7.71}$$

thus

$$f_0(x_k^{(-1)}) = 0, \quad 1 \le k \le L\text{-}1. \tag{7.72}$$

Therefore, for $j=-1$ and $1 \le k \le L\text{-}1$

$$f_{-1}(x_k^{(-1)}) + f_0(x_k^{(-1)}) = f_{-1}(x_k^{(-1)}) = \boldsymbol{I}_{V_0} f(x_k^{(-1)}) = f(x_k^{(-1)}), \tag{7.73}$$

and for $j=0$, $1 \le k \le L\text{-}1$

$$f_{-1}(x_k^{(0)}) + f_0(x_k^{(0)}) = \boldsymbol{I}_{V_0} f(x_k^{(0)}) + (\boldsymbol{f}_k^{(0)} - (\boldsymbol{I}_{V_0} f)_k^{(0)}) = f(x_k^{(0)}). \tag{7.74}$$

Equations (7.73) and (7.74) imply that function *f(x)* may be approximated at the next level of detail ($j = 0$) with

$$f(x) \simeq f_{-1}(x) + f_0(x). \tag{7.75}$$

Finally, for $1 \le j \le J-1$, generally it is possible to define

$$f_j(x) = \boldsymbol{I}_{W_j}(\boldsymbol{f}^{(j)} - (\mathcal{P}_{j-1} f)^{(j)}) = \sum_{k=-1}^{n_j-2} \hat{f}_{j,k} \psi_{j,k}(x), \tag{7.76}$$

where $(\mathcal{P}_{j-1} f)_k^{(j)} = \mathcal{P}_{j-1} f(x_k^{(j)})$, $-1 \le k \le n_j - 2$. It may be verified [CW96] that the function $f_{-1}(x)+f_0(x)+\ldots+f_{j-1}(x)$ interpolates the function *f(x)* on all collocation points $\left\{x_k^{(-1)}, 1 \le k \le L-1\right\}, \ldots, \left\{x_k^{(j)}, -1 \le k \le n_j - 2\right\}$. Specifically, setting $j = J$, $f_{-1}(x)+f_0(x)+\ldots+f_{J-1}(x) = \mathcal{P}_J f(x)$ and the interpolation conditions (7.62) hold. The

total cost of finding $\hat{f}$ is $8(L-1)+\sum_{j=0}^{J-1}(5j+9)n_j \le 6N \log N$ flops [CW96] where N is the total number of collocation points as defined in (7.65).

DWT for $\hat{f} \to f$

Assume that the function f is decomposed as it is proposed in (7.59) and that the coefficients $\hat{f}$ are obtained. Usually, some sort of processing then occurs and depending on the application, the coefficients may be changed (e.g. in denoising applications) or their number may be reduced (e.g. in data compression applications). In some cases, it is desirable to obtain values of the function f when the coefficients of the wavelet decomposition are known. This process is referred to as the *inverse DWT*. This direction of the DWT is straightforward and involves evaluating the expansion (7.59) at all collocation points $\left\{x_k^{(j)}\right\}, j \ge -1$. It takes $4(L-1)+\sum_{j=0}^{J-1}5jn_j \le 5N \log N$ flops to compute the vector $\boldsymbol{f}$ [CW96].

7.6. Summary

The 20th century has seen the introduction of wavelet theory with the potential for widespread application. The crucial property of wavelets is *the localization of the approximating function in time*. This answers the problem of the complete loss of localization information in the time domain that is characteristic of Fourier analysis, a powerful theory in every other aspect.

The Fourier Transform decomposes a signal in terms of complex exponential functions of different frequencies. With the FT it is possible to determine all the frequencies present in the signal but it is not possible to determine *when* they are present. In other words, a perfect frequency resolution is obtained but at the expense of no time resolution whatsoever. For non-stationary signals, time-domain information is necessary, hence a window function of a finite length was introduced and a *fixed resolution* in the time domain was obtained through the Short Term Fourier Transform (STFT). For complex signals, choosing a constant length window can produce some misleading results. Therefore, Wavelet Theory was introduced. By using a *fully scalable modulated window,* the WT gives *variable resolution* at all times, while the STFT can only give fixed resolution determined by the chosen support of the window used. The

important thing to note is that it is not possible to exactly know *what frequency exists* at *what time instance.* Rather it is possible only to know what **frequency bands** exist at what **time intervals**.

Based on the material presented in this Chapter, the advantages of wavelet theory compared to any other decomposition theory are:

- The wavelet theory is typically valid for much larger classes of functions than, for example, Taylor or Fourier theory;
- Wavelet coefficients convey *very detailed information* about f due to the structure of the chosen index set $\mathcal{J}$ which comprises information on scale and spatial location. This property of *localization in both the time and frequency* (scale) domain is the most important property of wavelets and which makes them so appealing for the analysis of transient signals.
- Unlike the Fourier Transform, where sinusoids are chosen as basis functions, and then the properties of the resulting expansion are examined, for wavelet analysis, the desired properties determine the resulting basis function.
- A new concept of *MultiResolution Analysis (MRA)* emerged from Wavelet Theory. In MRA, a signal is decomposed in terms of the resolution of detail.
- Wavelet-based analysis can naturally be applied on a digital computer with its basis functions defined by summations instead of integrals or derivatives.

There are many types of wavelets and wavelet families in use today. However, they all have some common characteristics. The two most important wavelet properties are *admissibility*, which stipulates that the wavelet function $\psi(t)$ must have a *band-pass spectrum* and *be a wave*, and the *regularity conditions* that state that the wavelet transform should decrease quickly with decreasing scale parameter s due to the fact that the wavelet has *n vanishing moments*.

For computation purposes, the Discrete Wavelet Transform (DWT) algorithm was developed to overcome the three drawbacks of the Continuous Wavelet Transform (CWT). Redundancy has been removed by introducing discrete wavelets and designing them to form an orthogonal basis. The problem of the infinite number of discrete wavelets needed in wavelet theory has been solved by introducing the scaling functions. The filter bank has solved the problem of the non-existence of analytical solutions of the CWT.

Finally, a detailed description of a wavelet-like system for solving initial boundary value problems, as proposed by Cai and Wang, is presented. This wavelet system will be used as a basis for developing a new and efficient wavelet-based technique for simulating non-linear electronic circuits subjected to high-frequency envelope modulated signals. The rationale for choosing this system lies in the fact that it introduces spline interpolation functions into the wavelet representation to deal with strong non-linearities. Although this wavelet does not have vanishing moments as is usual for wavelet bases, the point value vanishing property of this wavelet-like system enables a MRA formulation and a fast DWT calculation algorithm.

CHAPTER 8

A Novel Wavelet-based Approach for Transient Envelope Simulation

RF modulated signals comprising of a high-frequency carrier and a low-frequency information part represent a serious challenge for circuit simulators. The built-in ODE solver has to use very small time step in order to resolve the high-frequency carrier. However, it is the low-frequency part that has the useful information. Consequently, such signals result in lengthy simulations and the use of significant memory resources.

In this chapter, a novel approach for the simulation of non-linear circuits subject to modulated signals is presented. The approach combines the wavelet-based collocation technique presented in Section 7.5 with a multi-time method. The resulting novel simulation technique enables the desired trade-off between the required accuracy and computational efficiency. In addition, a non-linear model order reduction (MOR) technique is then applied with the aim to further improve computational efficiency.

8.1. Introduction

The two most common circuit simulation techniques that are employed in RF and microwave circuit simulators are Harmonic Balance [KS86], [LMA+97], [NV76] and Time-Domain Integration [N75]. Harmonic Balance is employed for periodic or quasi-periodic steady-state analysis of mildly non-linear circuits and hence can prove limited for the complex modulation formats encountered in today's high-speed systems or for systems involving strong non-linearities. Time-Domain Integration, on the other hand, is only suitable for baseband systems. For the simulation of circuits with digitally modulated high-frequency carriers with long bit sequences, Time-Domain Integration is excessively slow. As a result, there is a need for some form of general-purpose technique that can simulate over a long interval (e.g. to enable bit error ratio (BER) calculations) non-linear systems with transient high-frequency signals or complex modulated RF carriers.

To this end, several envelope transient analysis approaches have been proposed whereby a mixed-mode technique is implemented [NL96], [S96]. The essence of these approaches is that the envelope of a signal is treated by Time-Domain Integration and that the carrier is treated by Harmonic Balance. However, existing techniques have limitations, for example, restrictions in the bandwidth of the excitation signal [NL96] and the limitations of harmonic balance with respect to strong nonlinearities. Roychowdhury in [R01] proposes converting the differential-algebraic equations that describe the circuit to multi-time partial differential equations and applying time-domain methods directly to solve the resultant equations. Pedro [PC02] also employs a multi-time partial differential equation approach but uses a combination of Harmonic Balance and Time-Domain Integration to solve the resultant system.

The method proposed here follows Roychowdhury's approach in converting the differential-algebraic equations that describe the circuit to multi-time partial differential equations. In contrast to Roychowdhury's approach, the resultant equations are not solved directly. Instead, a wavelet-based collocation scheme is employed.

8.2. Multi-time partial differential equation (MPDE) approach

Consider a signal $x(t)$ that is composed of a high-frequency carrier modulated by a lower-frequency envelope where the envelope signal is assumed to be uncorrelated with the carrier. In this case, the signal may be represented in two independent time variables as follows:

$$x(t) = \bar{x}(t_1, t_2) \tag{8.1}$$

Time t_1 relates to the low-frequency envelope and t_2 relates to the high-frequency carrier. Now, consider a general nonlinear circuit described by:

$$\dot{q}\big(x(t)\big) = f\big(x(t)\big) + b(t), \tag{8.2}$$

where $b(t)$ is the input to the circuit and f is a linear or nonlinear function modelling resistive elements. q models the memoryless linear or nonlinear charges or fluxes. $x(t)$ is the state variable.

The corresponding multi-time partial differential equation (MPDE) system can be written [R01] as:

$$\frac{\partial q\big(\bar{x}(t_1,t_2)\big)}{\partial t_1} + \frac{\partial q\big(\bar{x}(t_1,t_2)\big)}{\partial t_2} = f\big(\bar{x}(t_1,t_2)\big) + b(t_1,t_2). \tag{8.3}$$

This multi-time partial differential equation can be solved using entirely time-domain approaches as employed by Roychowdhury [R01] or using a combination of time-step integration for the envelope and Harmonic Balance for the carrier as in [PC02]. However, for strongly non-linear circuits, the use of Harmonic Balance for the inner loop can prove limited as it leads to excessive computation. To overcome this limitation, a novel approach is proposed in this thesis. In essence, the MPDE system (8.3) is solved using a pseudo-wavelet collocation method derived from that proposed by Cai and Wang [CW96] as described in following section.

8.3. Wavelet collocation method for non-linear PDE

As just described, the MPDE approach yields a PDE representation of the system. A wavelet-based collocation method is then utilised for the purposes of solving the PDE system. The particular solution technique employed is that described in Chapter 7. The advantage of this approach is that it permits an adaptive multiresolution solution, which is ideal for simulating responses of envelope-modulated circuits. Christoffersen and Steer [CS01] also employ a version of the cubic spline collocation method presented in [CW96] but they apply it directly to the original ODE system rather than to a MPDE representation of the system as proposed in this thesis.

8.3.1. The rationale for choosing the wavelet basis

The signals considered in transient envelope analysis are signals with widely separated rates of variation. For example, such signals arise in communication circuits when a high-frequency RF carrier is modulated by a low-frequency information signal. The existence of a high-frequency component stipulates the use of a very short time step in existing time-domain simulation tools thus making such simulations excessively slow. On the other hand, it is the low-frequency information component that is of interest when analysing envelope-modulated signals. *If a high-frequency component can be resolved prior to the simulation of a low-frequency signal then the use of a significantly longer time-step would be possible, thereby greatly speeding up calculations.*

Thus, in this thesis, the *use of wavelet scheme* over some interval of interest is proposed in order *to resolve the contribution of high-frequency components* yielding a semidiscretised wavelet collocation representation with respect to t_2. In effect the

MPDE system representation is translated to an ODE system with respect to t_1. To obtain a fully discretised wavelet collocation method, the time-derivative with respect to t_1 (representing slowly-varying envelope) is replaced by a suitable difference equation. An adaptive Backward-Euler predictor-corrector approach may then be employed to obtain the final solution.

8.3.2. The wavelet basis and collocation points

Let $[t_0, t_{end}]$ be the interval of interest in the time domain. The technique proposed in this thesis involves approximating the unknown function $\bar{x}(t_1,t_2)$ with a wavelet series $\bar{x}_J(t_1,t_2)$ in the t_2 dimension, i.e.

$$\begin{aligned}\bar{x}_J(t_1,t_2) = {} & \hat{x}_{-1,-3}(t_1)\eta_1(t_2)+\hat{x}_{-1,-2}(t_1)\eta_2(t_2)+\hat{x}_{-1,-1}(t_1)\varphi_b(t_2)+\sum_{k=0}^{L-4}\hat{x}_{-1,k}(t_1)\varphi_k(t_2) \\ & +\hat{x}_{-1,L-3}(t_1)\varphi_b(L-t_2)+\sum_{j=0}^{J-1}\sum_{k=-1}^{n_j-2}\hat{x}_{j,k}(t_1)\psi_{j,k}(t_2)+\hat{x}_{-1,L-2}(t_1)\eta_2(L-t_2) \\ & +\hat{x}_{-1,L-1}(t_1)\eta_1(L-t_2)\end{aligned} \tag{8.4}$$

where the integer $J \geq 0$ determines the maximum wavelet level being considered. The parameter $L > 4$ determines the interval $[0, L]$ which uniquely corresponds to the initial interval $[t_0, t_{end}]$ and $t_2 \in [0, L]$. $\varphi(t)$ and $\varphi_b(t)$ are the interior and boundary scaling functions respectively given in (7.29) and (7.30). $\psi(t)$ and $\psi_b(t)$ are the interior and boundary wavelet functions respectively given in (7.36) and (7.37). $\eta_1(t)$ and $\eta_2(t)$ are the spline functions introduced to approximate boundary-nonhomogeneities as described in (7.45). A detailed description and properties of the aforementioned functions is given in Chapter 7.

$\hat{x}(t_1)$ are the unknown coefficients which are a function of t_1 only. The total number of unknown coefficients is $N = 2^J L + 3$ where J determines the level of wavelet coefficients taken into account when approximating $\bar{x}(t_1,t_2)$. Note that the total number of coefficients in this instance is four more from the one stated in (7.65). This is due to the fact that the interpolating spline function $\boldsymbol{I}_{b,j}f(x)$ coefficients are also taken into account.

For the purposes of clarity, denote:

$$\bar{x}_J(t_1,t_2) = \sum_{k=1}^{N}\hat{x}_k(t_1)\Psi_k(t_2) \tag{8.5}$$

From this point forward, $\Psi_k(t)$ shall be referred to as "wavelets" where it is understood that these comprise the scaling functions, $\varphi(t)$, the wavelet functions, $\psi(t)$ and the nonhomogeneity functions, $\eta(t)$.

The collocation points chosen are as given in [CW96] and are:

$$\begin{aligned} &t_{2,1}^{(-1)} = 0; \quad t_{2,2}^{(-1)} = 0.5; \quad t_{2,k}^{(-1)} = k-2, 3 \le k \le L+1; \\ &t_{2,L+2}^{(-1)} = L-0.5; \quad t_{2,L+3}^{(-1)} = L \\ &t_{2,-1}^{(j)} = \frac{1}{2^{j+2}}; \quad t_{2,k}^{(j)} = \frac{k+1.5}{2^j}, \; 0 \le k \le n_j - 3; \; t_{2,n_j-2}^{(j)} = L - \frac{1}{2^{j+2}} \end{aligned} \quad , \qquad (8.6)$$

where again $n_j = 2^j L$. Subscript *2* refers to the t_2 variable. Equation (8.3) is then collocated on collocation points to result in a semidiscretised wavelet collocation representation.

8.3.3. Wavelet collocation method

The set of ordinary differential equations (8.2) is first written as a set of multi-time partial differential equations (8.3) as suggested in Section 8.2. Note that t_1 relates to the low-frequency envelope and t_2 relates to the high-frequency carrier. Equation (8.3) is then collocated on collocation points (8.6) to result in a semidiscretised wavelet collocation method. To obtain a fully discretised wavelet collocation method, the time-derivative with respect to t_1 (representing the slowly-varying envelope) is replaced by a suitable difference equation. An adaptive Backward-Euler predictor corrector approach is then employed in contrast to a simple Forward Euler that was suggested in [CW96]. This leads to significant gains in efficiency compared to fixed-step approaches. Consequently, the overall technique can be implemented in an efficient manner. It obviates the need for solving non-linear algebraic equations at each timestep thereby removing the potential difficulties that arise in other simulation approaches when large-scale non-linear systems are present.

Furthermore, in most cases, many of the wavelet coefficients may be neglected within a given tolerance ε [CW96]. This permits the number of wavelet functions included to be adjusted dynamically thereby reducing the computing requirements while at the same time achieving a satisfactory level of accuracy. For example, if

$$\left|\hat{x}_{j,k}(t_1)\right| < \varepsilon$$

then the wavelet function associated with this coefficient may be neglected. Furthermore, if the maximum coefficient in any level of resolution, *J*, is less than the

tolerance, then the level J of the wavelet expansion can be decreased to reduce computational requirements, i.e. $\max|\hat{x}_{j,k}(t_l)| < \varepsilon$ would imply decreasing J to J' in (8.4) where $J'<J$.

8.4. Numerical results for sample systems

The full wavelet approach described in this Section has been tested on two sample non-linear systems, a diode rectifier circuit shown in Fig. 8.1 and a MESFET amplifier given in Fig. 8.5. The complete parameters, details and equations for these sample circuits are given in the Appendix G.

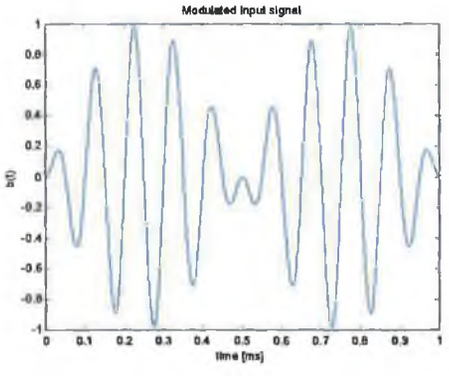


Fig. 8.1. *Modulated input signal*

Both systems are excited with an excitation signal of the form:

$$b(t) = sin(\frac{2\pi}{T_1}t)sin(\frac{2\pi}{T_2}t), \tag{8.7}$$

where T_1 corresponds to the envelope period (slower varying signal) and T_2 corresponds to the carrier period (faster varying signal). Fig. 8.1. shows the excitation signal for $T_1 = 1ms$ and $T_2 = 0.1ms$. It is clear that the expression from (8.7) represents a slowly changing sinusoid corresponding to T_1 modulated by a fast changing sinusoid corresponding to T_2.

8.4.1. Non-linear diode rectifier circuit

The sample non-linear diode rectifier circuit is given in Fig. 8.2. The rectifier is excited with the input signal given in (8.7) with $T_1 = 1ms$ and $T_2 = 0.1ms$.

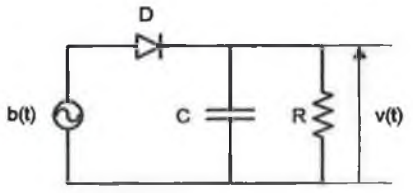


Fig. 8.2. *Diode rectifier circuit*

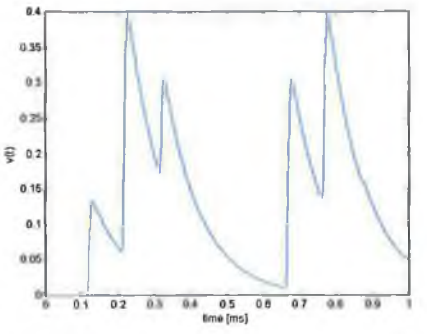


Fig. 8.3. *Result from ODE solver with a very short timestep*

Fig. 8.3 shows the output from a commercial ordinary differential equation solver with a very short time step in order to obtain a highly accurate version of the output voltage to act as a benchmark for the purposes of confirming the accuracy of the proposed new simulation technique. Fig. 8.4 shows a result with a very coarse level of resolution (*J=0, L=80*) , i.e. only scaling functions are utilised in the representation of the unknown voltage in (8.4). It is clear that the salient behaviour of the response is successfully captured.

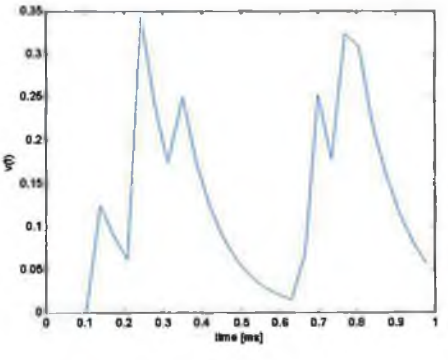


Fig. 8.4. *Result with a very coarse level of resolution (J=0, L=80)*

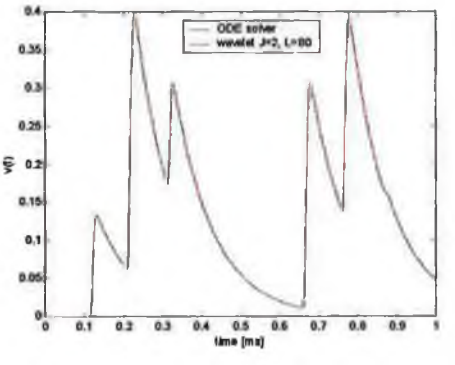


Fig. 8.5. *Sample result from new method*

In order to improve the accuracy of the response, two wavelet levels are added to the representation in (8.4) and Fig. 8.5 shows the output voltage at this new level of resolution (*J=2, L=80*). As evidenced by this result, the new method achieves a good level of accuracy. Obviously, greater accuracy can be achieved by increasing the level of resolution in the wavelet scheme (or by setting a tighter tolerance value) but at the cost of increasing simulation time.

The non-linear diode rectifier circuit given in Fig. 8.2 is deliberately selected as it is strongly non-linear in nature as can be seen from Fig. 8.3. The ability to efficiently simulate the behaviour of this circuit with good accuracy provides a strong

recommendation for employing the wavelet-based simulation technique presented here to simulate highly non-linear circuits subjected to input signals that have widely separated rates of variation.

8.4.2. MESFET amplifier

The second example taken is that of the single-ended practical MESFET amplifier shown in Fig. 8.6. The amplifier is described by ten non-linear differential equations that are stiff in nature. The equations and MESFET parameters are given in Appendix G. The input to the circuit is a 2GHz wave modulated by a 0.2GHz wave.

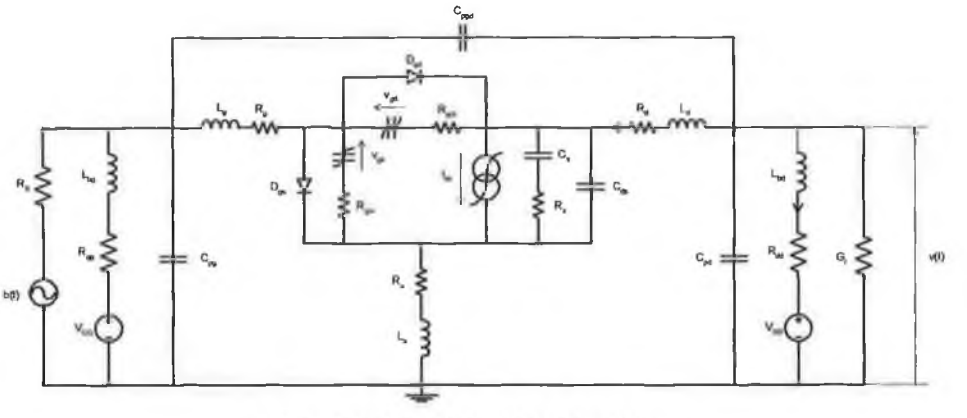


Fig. 8.6. *Simple MESFET Amplifier*

Fig. 8.7 shows the output voltage obtained when a fourth-order Adams-Moulton predictor-corrector technique is employed with a time-step of 0.1ps. This is deemed an accurate representation of the output voltage for comparative purposes.

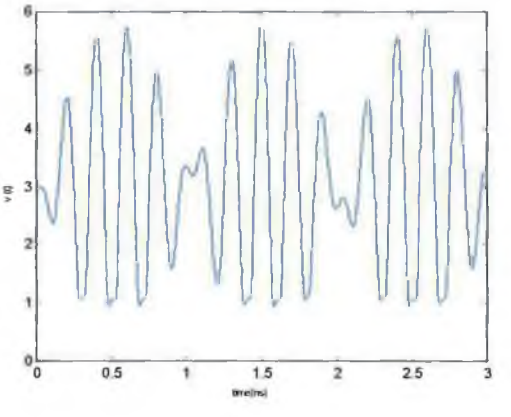


Fig 8.7. *Result with Adams-Moulton technique*

Fig. 8.8 shows a result when the novel technique with a very coarse resolution (*J=1, L=80*) is employed. As can be seen, the general nature of the circuit response is

obtained. However, due to the complex structure of the circuit, the lower-order wavelet approximation (*J=1*) is not sufficient in this case to acquire the fine details of the output. Hence, there is a need to use a higher-order wavelet approximation (*J=2, L=80*) as shown in Fig. 8.19. It is clear that a high degree of accuracy is achieved. As evidenced by this result, the technique is highly effective in predicting the output voltage for structurally complex non-linear circuits.

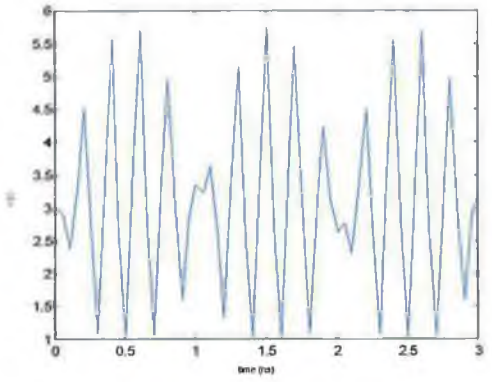


Fig. 8.8. *Output voltage with a coarse level of resolution*

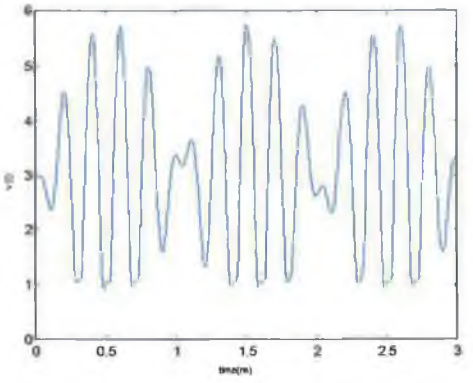


Fig. 8.9. *Output voltage with a fine level of resolution*

These results are published in [CD03]. They show that the proposed *full* wavelet-based technique is capable of accurately capturing the transient response of a non-linear circuit excited with an envelope modulated signal even at a very coarse level of resolution. In the following section, an extension to the described technique is presented. The aim of this extension is to further increase the efficiency of the technique by employing a *non-linear* model order reduction.

8.5. Wavelet collocation method in conjunction with MOR

The wavelet collocation scheme for non-linear PDEs proposed in the previous section has great flexibility when it comes to obtaining a result of a certain required accuracy. In practice, accuracy is simply determined by the chosen wavelet level *J* and can be dynamically adjusted during the calculation process. However, the drawback of the presented scheme is that it results in a large system of ODEs that needs to be solved. This can be very costly in terms of computational time and resources.

Therefore, to address this issue, this section presents a modification of the wavelet-based collocation approach presented in the Section 8.3. This approach is greatly enhanced in that a non-linear model reduction strategy similar to that in [GN99]

is employed within the proposed envelope simulation technique to obtain very high efficiencies.

As will be shown by results, this dramatically improves the efficiency of calculation and drastically reduces computational requirements but without a complementary loss in accuracy.

8.5.1. Matrix representation of full wavelet collocation scheme

Consider the non-linear circuit equation described in the standard form of a non-linear ordinary differential equation:

$$\frac{dx(t)}{dt} + cx(t) + f(x(t)) = b(t) \tag{8.8}$$

where c is constant relating to the linear part of the circuit, f describes the circuit non-linearity and b is the excitation signal. Following the MPDE approach, equation (8.8) may be written as:

$$\frac{\partial \bar{x}(t_1,t_2)}{\partial t_1} + \frac{\partial \bar{x}(t_1,t_2)}{\partial t_2} + c\bar{x}(t_1,t_2) + f(\bar{x}(t_1,t_2)) = b(t_1,t_2) \tag{8.9}$$

Now, the unknown $x(t_1,t_2)$ may be approximated with $\bar{x}_J(t_1,t_2)$ from equation (8.5), i.e.

$$\bar{x}(t_1,t_2) = \bar{x}_J(t_1,t_2) = \sum_{k=1}^{N} \hat{x}_k(t_1)\Psi_k(t_2) \tag{8.10}$$

Then, the expression in (8.10), if written for all collocation points in t_2, may be expressed as follows at a specific point in time t_1:

$$\bar{\boldsymbol{x}}_{JN}(t_1) = \boldsymbol{E}\hat{\boldsymbol{x}}(t_1) \tag{8.11}$$

where $\boldsymbol{E}$ is a constant N-dimensional square matrix whose columns comprise the values of the N wavelet functions, $\Psi_k(t_2)$, at N collocation points:

$$\boldsymbol{E}(t_2) = \begin{bmatrix} \Psi_1(t_2^1) & \Psi_2(t_2^1) & \cdots & \Psi_N(t_2^1) \\ \Psi_1(t_2^2) & \Psi_2(t_2^2) & \cdots & \Psi_N(t_2^2) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_1(t_2^N) & \Psi_2(t_2^N) & \cdots & \Psi_N(t_2^N) \end{bmatrix} \tag{8.12}$$

where t_2^k, $k = 1,...,N$ denotes k^{th} collocation point. The matrix is evaluated once at the outset of the algorithm. $\bar{\boldsymbol{x}}_{JN}(t_1)$ is an N-dimensional column vector of the unknown state-variables and $\hat{\boldsymbol{x}}(t_1)$ is an N-dimensional column vector of the unknown wavelet coefficients at the collocation points in t_2 at a specific instant in t_1:

$$\hat{\boldsymbol{x}}(t_1)=\begin{bmatrix}\hat{x}_1(t_1)\\ \hat{x}_2(t_1)\\ \vdots\\ \hat{x}_N(t_1)\end{bmatrix} \tag{8.13}$$

Substitution of (8.10) and (8.11) into (8.9) yields:

$$\boldsymbol{E}\frac{d\hat{\boldsymbol{x}}}{dt_1}=-\boldsymbol{D}\hat{\boldsymbol{x}}+\boldsymbol{f}_N(\hat{\boldsymbol{x}})+\boldsymbol{b}_N \tag{8.14}$$

where $\boldsymbol{D}$ is an N dimensional matrix given in (8.15) whose columns are formed from the derivatives of the wavelet functions in (8.4) evaluated at each of the N collocation points in t_2.

$$\boldsymbol{D}(t_2)=\begin{bmatrix}\frac{d\Psi_1(t_2^1)}{dt_2}+c\Psi_1(t_2^1) & \frac{d\Psi_2(t_2^1)}{dt_2}+c\Psi_2(t_2^1) & \cdots & \frac{d\Psi_N(t_2^1)}{dt_2}+c\Psi_N(t_2^1)\\ \frac{d\Psi_1(t_2^2)}{dt_2}+c\Psi_1(t_2^2) & \frac{d\Psi_2(t_2^2)}{dt_2}+c\Psi_2(t_2^2) & \cdots & \frac{d\Psi_N(t_2^2)}{dt_2}+c\Psi_N(t_2^2)\\ \vdots & \vdots & \ddots & \vdots\\ \frac{d\Psi_1(t_2^N)}{dt_2}+c\Psi_1(t_2^N) & \frac{d\Psi_2(t_2^N)}{dt_2}+c\Psi_2(t_2^N) & \cdots & \frac{d\Psi_N(t_2^N)}{dt_2}+c\Psi_N(t_2^N)\end{bmatrix} \tag{8.15}$$

Again, $\boldsymbol{D}$ is evaluated only once at the outset of the algorithm. $\boldsymbol{f}_N$ and $\boldsymbol{b}_N$ are column vectors comprising the values of f and b at the collocation points as in:

$$\boldsymbol{f}_N=\begin{bmatrix}f(t_1,t_2^1)\\ f(t_1,t_2^2)\\ \vdots\\ f(t_1,t_2^N)\end{bmatrix},\qquad \boldsymbol{b}_N=\begin{bmatrix}b(t_1,t_2^1)\\ b(t_1,t_2^2)\\ \vdots\\ b(t_1,t_2^N)\end{bmatrix} \tag{8.16}$$

Thus, equation (8.14) represents an ordinary differential equation in the t_1 domain. To obtain a solution to this equation in an efficient manner, the model order reduction technique described in the next section is proposed.

8.5.2. Model order reduction technique

The crucial step introduced in this section is the application of a non-linear model reduction process within the proposed wavelet-based collocation scheme. As in the wavelet method proposed in Section 8.3, equation (8.3) is first collocated on collocation points (8.6) in the time-domain t_2 to result in a semidiscretised equation system (8.14). At this juncture, the technique differs significantly from that presented in

Section 8.3. Instead of directly solving for the unknown state-variables and output $y(t)$ at each time-step in t_1, a non-linear model reduction strategy is employed. The particular model reduction strategy chosen is based on that proposed by Gunupudi and Nakhla [GN99] and will be briefly described here.

First, the vector of coefficients, $\hat{\boldsymbol{x}}(t_1)$, is expanded in a Taylor series as follows:

$$\hat{\boldsymbol{x}}(t_1) = \sum_{i=0}^{\infty} \boldsymbol{a}_i (t_1 - t_1^0)^i \tag{8.17}$$

where t_1^0 is the initial time and where the coefficients, $\boldsymbol{a}_i$, may be computed recursively as in [GN99]. Then a Krylov space is formed for $\boldsymbol{a}_i$:

$$\boldsymbol{K} = [\boldsymbol{a}_0 \quad \boldsymbol{a}_1 \quad \cdots \quad \boldsymbol{a}_q], \tag{8.18}$$

where q is the order of the reduced system and is significantly less than the order of the original system N.

An orthogonal decomposition of $\boldsymbol{K}$ results in:

$$\boldsymbol{K} = \boldsymbol{Q}\boldsymbol{R}, \tag{8.19}$$

where $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_q$. $\boldsymbol{I}_q$ is the q dimensional identity matrix. $\boldsymbol{Q}$ is then employed to perform a congruent transformation of:

$$\hat{\boldsymbol{x}} = \boldsymbol{Q}\tilde{\boldsymbol{x}}. \tag{8.20}$$

where $\tilde{\boldsymbol{x}}$ is the q dimensional ($q<<N$) vector of new unknown coefficients. Consequently, a new reduced equation system is formed as:

$$\boldsymbol{Q}^T\boldsymbol{E}\boldsymbol{Q}\frac{d\tilde{\boldsymbol{x}}}{dt_1} = -\boldsymbol{Q}^T\boldsymbol{D}\boldsymbol{Q}\tilde{\boldsymbol{x}} + \boldsymbol{Q}^T\boldsymbol{f}_N(\boldsymbol{Q}\tilde{\boldsymbol{x}}) + \boldsymbol{Q}^T\boldsymbol{b}_N \tag{8.21}$$

or, in shorter notation,

$$\tilde{\boldsymbol{E}}\frac{d\tilde{\boldsymbol{x}}}{dt_1} = -\tilde{\boldsymbol{D}}\tilde{\boldsymbol{x}} + \boldsymbol{Q}^T\boldsymbol{f}_N(\boldsymbol{Q}\tilde{\boldsymbol{x}}) + \tilde{\boldsymbol{b}}_N \tag{8.22}$$

where

$$\tilde{\boldsymbol{E}} = \boldsymbol{Q}^T\boldsymbol{E}\boldsymbol{Q}, \tilde{\boldsymbol{D}} = \boldsymbol{Q}^T\boldsymbol{D}\boldsymbol{Q} \text{ and } \tilde{\boldsymbol{b}}_N = \boldsymbol{Q}^T\boldsymbol{b}_N. \tag{8.23}$$

Thus instead of solving an N^{th} order system at each time step to obtain the unknown state-variables and the output quantity $y(t)$, a reduced-order system (8.22) of transformed coefficients is solved. A trapezoidal-rule integration scheme is employed because of its superior stability qualities. After solving this new system, of dimension $q << N$, the values for $\tilde{\boldsymbol{x}}$ over the entire time domain of interest is determined. Once the q coefficients, $\tilde{\boldsymbol{x}}$, have been determined, $\hat{\boldsymbol{x}}(t_1)$ and consequently, $\bar{\boldsymbol{x}}_{JN}(t_1) = \boldsymbol{E}\hat{\boldsymbol{x}}(t_1)$,

may be obtained in one single post-processing step involving only matrix multiplication. The above solution process is thus significantly more efficient than solving directly for $\bar{x}_{JN}(t_l)$ at each time step as was done in Section 8.3.

8.6. Numerical results for sample systems

The same non-linear diode rectifier circuit and MESFET amplifier as in Section 8.4. are used to test the accuracy and the efficiency of the wavelet–based scheme with the applied nonlinear model-order reduction technique. The results, as published in [DCB04a] and [DCB05], will confirm that for a comparable computation time, significant gains in accuracy may be achieved by employing the proposed approach with model order reduction as opposed to simply using a lower-order full wavelet scheme.

8.6.1. Non-linear diode rectifier circuit

Consider, again the non-linear diode rectifier circuit given in Fig. 8.2. The output from a full wavelet scheme with no model order reduction applied, i.e. from the technique described in Section 8.3, is presented in Fig. 8.10. For the chosen wavelet parameters *J=1* and *L=80,* the size of the ODE system is *N=163*. An adaptive Backward-Euler predictor corrector approach is employed for obtaining the solution. Good agreement is achieved when compared to the 'accurate' result given in Fig. 8.3. However, significant computer resources are required to solve an ODE system involving 163 unknown variables.

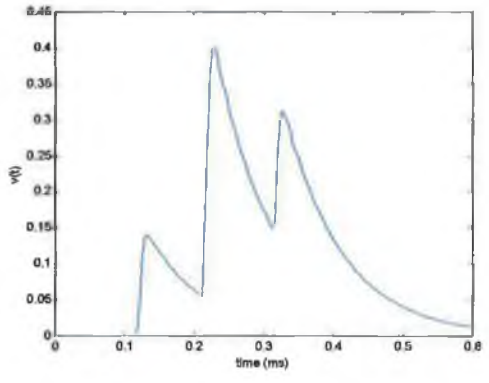


Fig. 8.10. *Result from a full wavelet scheme (J=1, L=80)*

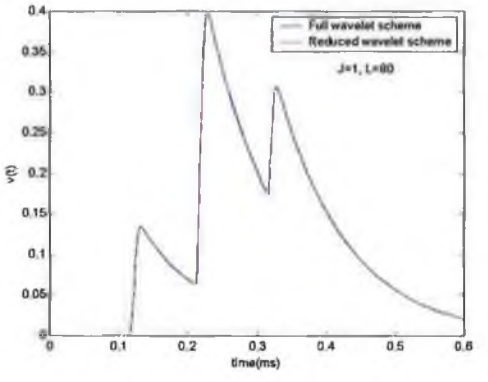


Fig. 8.11. *Result from wavelet scheme (J=1, L=80) with MOR applied (q=5)*

Fig. 8.11. shows the output when the model-order reduction technique proposed in this Section is applied. For the same wavelet parameters (*J=1, L=80*), the initial system of *N=163* unknown wavelet coefficients is reduced to *q=5* before obtaining the solution for the reduced-order system (8.22). In terms of accuracy, the relative difference between the result from the full wavelet scheme and the results obtained having applied the model reduction technique is negligible. However, in terms of computation time, the result obtained with the model reduction technique is computed in only *7%* of the time necessary for the full wavelet scheme. This excellent gain in computational efficiency is due to the fact that instead of solving an ODE system with 163 unknowns, a system with only 5 unknowns is solved at each time step.

Finally, Fig. 8.12 shows the result when a lower order full wavelet scheme is employed. In this case, $L = 5$ and $J = 0$ in (8.4). This results in an $N=8^{th}$ order system of equations which has similar computational requirements to the reduced wavelet scheme with $q = 5$.

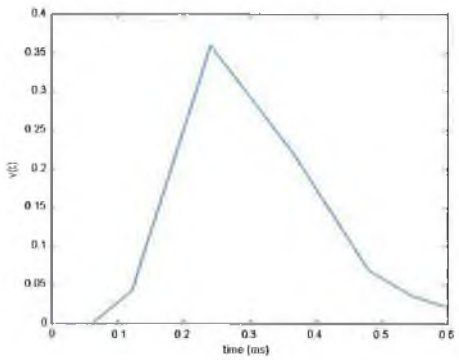


Fig. 8.12. *Result with lower-order full wavelet scheme (J=0, L=5)*

As can be seen from Fig. 8.12, there is a significant loss in accuracy. This result clearly confirms that the approach presented in this section is significantly better than simply employing a full lower-order wavelet scheme especially when circumstances require high computational efficiency.

8.6.2. MESFET amplifier

Fig. 8.13 presents the MESFET response when the full wavelet scheme (*J=2, L=80*) is employed. The size of the resultant ODE system is *N=323*.

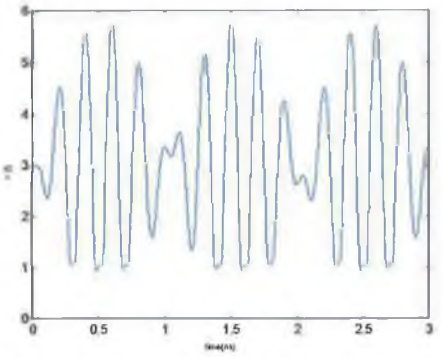


Fig. 8.13. *Result with full wavelet scheme (J=2,L=80)*

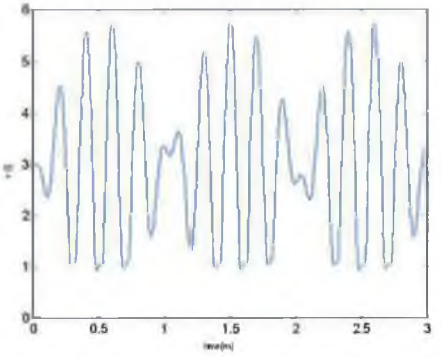


Fig. 8.14. *Result from a wavelet scheme (J=2, L=80) with MOR applied (q=20)*

Fig. 8.14 shows the MESFET output when model order reduction ($q=20$) has been applied. This result obtained with the model reduction technique is computed in only *11%* of the time necessary for the full wavelet scheme. Again, applying the MOR technique has resulted in vast gains in terms of computational efficiency when calculating the response of a complex electronic circuit.

8.7. Conclusion

In this Chapter, a novel approach for the simulation of high-frequency non-linear circuits subject to signals with widely separated rates of variation, i.e. envelope modulated signals, is presented. The proposed approach combines a wavelet-based collocation technique with a multi-time approach to result in a novel simulation technique, which enables the desired trade-off between the required accuracy and computational efficiency. A non-linear model-order reduction technique is then applied with the aim to further improve computational efficiency.

Two sample systems have illustrated the efficacy and the accuracy of the proposed envelope simulation technique. The results for the diode rectifier response confirm the efficacy of the proposed method for non-linear circuits, while the simulation results for the MESFET amplifier response confirm the efficacy of the proposed method for stiff complex non-linear circuits.

The principal advantage of the proposed method is that it may be applied in the case of strongly non-linear complex circuits and that it permits an effective trade-off between accuracy and speed.

CHAPTER 9

An Efficient Non-linear Circuit Simulation Technique for IC Design

In the initial stage of a design cycle, the circuit designer is interested in the overall functional behaviour of the designed circuit, i.e. will the integrity of the desired logical states be preserved at the output? In order to ascertain this, the designer needs to perform numerous simulations before settling on a final design. Any change in the requirements for the circuit design will necessitate the simulation process to restart from the beginning. However, the complexity of today's integrated circuits is such that these simulations are computationally expensive both in terms of time and computer resources. The overall result is a prolonged design cycle that is economically unacceptable. Hence, there is a need for a simulation technique that enables the designer to obtain the circuit response with the desired accuracy and within a reasonable time-frame. Ideally, the first phase of the design process should involve obtaining a rough initial result for the circuit response to verify the functionality of the design. In the second phase, when a higher degree of accuracy for fine-tuning the designed IC is sought, the possibility of reusing results from the first phase would yield huge gains in the efficiency of a simulation, thereby leading to major savings in the design time and ultimately reducing the cost of the designed IC.

Based on the approach presented in Chapter 8, a novel wavelet-based method for the analysis and simulation of IC circuits with the potential to greatly shorten the IC design cycle is presented in this chapter. The efficiency of the proposed method has been further improved using a model order reduction technique to obtain even more gains in terms of computational speed.

9.1. Formation of an approximation with a higher-degree of accuracy from an available lower-degree accuracy approximation

Assume that a preliminary circuit response is obtained by applying the technique presented in Chapter 8. If now, a response with a higher degree of accuracy is required, the wavelet series approximating the unknown function, $\bar{x}(t_1, t_2)$, can be expanded for

another layer, i.e.

$$\bar{x}_{J_1}(t_1,t_2)=\sum_{k=1}^{N_1}\hat{x}_k(t_1)\Psi_k(t_2), \tag{9.1}$$

where $J_1=J+1$ and the total number of unknown coefficients is now $N_1=2^{J_1}L+3$. At this point, two options are available.

Firstly, the method proposed in Chapter 8 can be implemented from scratch to obtain the circuit response. The size of ODE system to be solved is increased from $N=2^{J}L+3$ to $N_1=2^{J_1}L+3=2^{J+1}L+3$ and consequently, the computational requirements for obtaining the required solution are also increased.

Alternatively, the following approach may be applied to obtain the circuit response with increased accuracy. First, write (9.1) as:

$$\bar{x}_{J_1}(t_1,t_2)=\sum_{k=1}^{N_1}\hat{x}_k(t_1)\Psi_k(t_2)=\sum_{k=1}^{N}\hat{x}_k(t_1)\Psi_k(t_2)+\sum_{k=N+1}^{N_1}\hat{x}_k(t_1)\Psi_k(t_2) \tag{9.2}$$

or, after setting $M=N_1-N=2^{J}L$, the wavelet series approximating the unknown function $\bar{x}(t_1,t_2)$ can be written as:

$$\bar{x}_{J_1}(t_1,t_2)=\sum_{k=1}^{N}\hat{x}_k(t_1)\Psi_k(t_2)+\sum_{m=1}^{M}\hat{x}_{N+m}(t_1)\Psi_{N+m}(t_2). \tag{9.3}$$

The first term in (9.3) depends solely on coefficients from previous layers. The values for these coefficients at the collocation points up to the layer J are already known from previous calculations and any additional required values can be obtained using a standard interpolation technique [ML91]. The second term in (9.3) consists solely of unknown coefficients from the added layer, and thus, they need to be calculated.

Now, for presentation purposes, consider the following notation:

$$\hat{x}_k(t_1)=\hat{c}_k(t_1),\quad k=1,\ldots,N \tag{9.4}$$

and

$$\hat{x}_k(t_1)=\hat{g}_m(t_1),\quad k=N+1,\ldots,N+M;\quad m=1,\ldots,M. \tag{9.5}$$

Thus, the wavelet series approximating the unknown function, $\bar{x}(t_1,t_2)$, can be written as:

$$\bar{x}_{J_1}(t_1,t_2)=\sum_{k=1}^{N}\hat{c}_k(t_1)\Psi_k(t_2)+\sum_{m=1}^{M}\hat{g}_m(t_1)\Psi_{N+m}(t_2) \tag{9.6}$$

The expression in (9.6), if written for the M collocation points of the added layer in t_2, may be expressed as follows at a specific point in time t_1:

$$\bar{\boldsymbol{x}}_{J_1M}(t_1)=\boldsymbol{E}_0\hat{\boldsymbol{c}}(t_1)+\boldsymbol{E}_1\hat{\boldsymbol{g}}(t_1) \tag{9.7}$$

where $\hat{\boldsymbol{g}}(t_1)$ is an M-dimensional column vector of the *unknown* wavelet coefficients of

layer J_1. $\hat{\boldsymbol{c}}(t_1)$ is an N-dimensional column vector of the *known* wavelet coefficients at the collocation points in t_2 at a specific instant in t_1 and its entries are either already known directly or may be obtained as interpolated values for any time t_1. $\boldsymbol{E_0}$ is a constant $M\times N$-dimensional matrix whose columns comprise the values of the N wavelet functions, $\Psi_k(t_2)$, at the M collocation points of the extra layer, while $\boldsymbol{E_1}$ is a constant M-dimensional square matrix with $\Psi_k(t_2)$, at the M collocation points of the extra layer as its entries. All constant matrices are evaluated only once at the outset of the algorithm. $\bar{\boldsymbol{x}}_{J_1 M}(t_1)$ is an M-dimensional column vector of the unknown state-variables on layer J_1.

Substitution of (9.6) and (9.7) into (8.3) yields:

$$\boldsymbol{E}_1 \frac{d\hat{\boldsymbol{g}}}{dt_1} = -\boldsymbol{D}_1\hat{\boldsymbol{g}} - \boldsymbol{E}_0 \frac{d\hat{\boldsymbol{c}}}{dt_1} - \boldsymbol{D}_0\hat{\boldsymbol{c}} + \boldsymbol{f}_M(\hat{\boldsymbol{c}},\hat{\boldsymbol{g}}) + \boldsymbol{b}_M \tag{9.8}$$

where $\boldsymbol{D_0}$ is an $M\times N$ dimensional matrix whose columns are formed from the derivatives of the wavelet functions evaluated at each of the M collocation points of the extra layer and $\boldsymbol{D_1}$ is an $M\times M$ dimensional matrix analogous to matrix $\boldsymbol{D}$ in (8.15). Again, $\boldsymbol{D_0}$ and $\boldsymbol{D_1}$ are evaluated only once at the outset of the algorithm. $\boldsymbol{f_M}$ and $\boldsymbol{b_M}$ are column vectors comprising the values of f and b at the collocation points of level J_1.

Bearing in mind the notation introduced in (9.4) and (9.5), $\frac{d\hat{\boldsymbol{c}}}{dt_1}$ may be expressed, using (8.9), as a function of $\hat{\boldsymbol{c}}$:

$$\frac{d\hat{\boldsymbol{c}}}{dt_1} = \boldsymbol{E}^{-1}[-\boldsymbol{D}\hat{\boldsymbol{c}} + \boldsymbol{f}_N(\hat{\boldsymbol{c}}) + \boldsymbol{b}_N] \tag{9.9}$$

Substituting (9.9) in (9.8) yields the following equation:

$$\boldsymbol{E}_1 \frac{d\hat{\boldsymbol{g}}}{dt_1} = -\boldsymbol{D}_1\hat{\boldsymbol{g}} + (\boldsymbol{E}_0\boldsymbol{E}^{-1}\boldsymbol{D} - \boldsymbol{D}_0)\hat{\boldsymbol{c}} + \boldsymbol{f}_M(\hat{\boldsymbol{c}},\hat{\boldsymbol{g}}) - \boldsymbol{E}_0\boldsymbol{E}^{-1}\boldsymbol{f}_N(\hat{\boldsymbol{c}}) + \boldsymbol{b}_M - \boldsymbol{E}_0\boldsymbol{E}^{-1}\boldsymbol{b}_N. \tag{9.10}$$

This may be written for convenience as:

$$\boldsymbol{E}_1 \frac{d\hat{\boldsymbol{g}}}{dt_1} = -\boldsymbol{D}_1\hat{\boldsymbol{g}} + \boldsymbol{F}_M(\hat{\boldsymbol{c}},\hat{\boldsymbol{g}}) + \boldsymbol{B}_M \tag{9.11}$$

where

$$\boldsymbol{F}_M(\hat{\boldsymbol{c}},\hat{\boldsymbol{g}}) = (\boldsymbol{E}_0\boldsymbol{E}^{-1}\boldsymbol{D} - \boldsymbol{D}_0)\hat{\boldsymbol{c}} + \boldsymbol{f}_M(\hat{\boldsymbol{c}},\hat{\boldsymbol{g}}) - \boldsymbol{E}_0\boldsymbol{E}^{-1}\boldsymbol{f}_N(\hat{\boldsymbol{c}}) \tag{9.12}$$

and

$$\boldsymbol{B}_M = \boldsymbol{b}_M - \boldsymbol{E}_0\boldsymbol{E}^{-1}\boldsymbol{b}_N. \tag{9.13}$$

Equation (9.11) represents a $M\times M$ system of ordinary differential equations where the unknowns $\hat{\boldsymbol{g}}$ may be readily determined using a standard numerical technique for

solving a system of ordinary differential equations [ML91]. The system in (9.11) is significantly smaller in dimension than that in (8.9) in that it involves only M unknowns rather than $N+M$ unknowns when written for the same wavelet approximation level $J+1$. Therefore, the computational cost in obtaining the circuit response is significantly reduced.

9.2. Numerical results of sample systems

The proposed method is tested on the sample circuits from Section 8.4: a diode rectifier given in Fig. 8.2 and a MESFET amplifier given in Fig. 8.6. The results were reported in [DCB05].

9.2.1. Non-linear diode rectifier circuit

To emphasize the gains in accuracy achieved by the addition of an extra layer in the wavelet approximation series, Fig. 9.1. shows an example with wavelet layers $J=1$ and $J=2$. The collocation points range parameter, L, was deliberately chosen to be very low $(L=10)$ so that gains in the accuracy due to adding an extra layer would be highlighted.

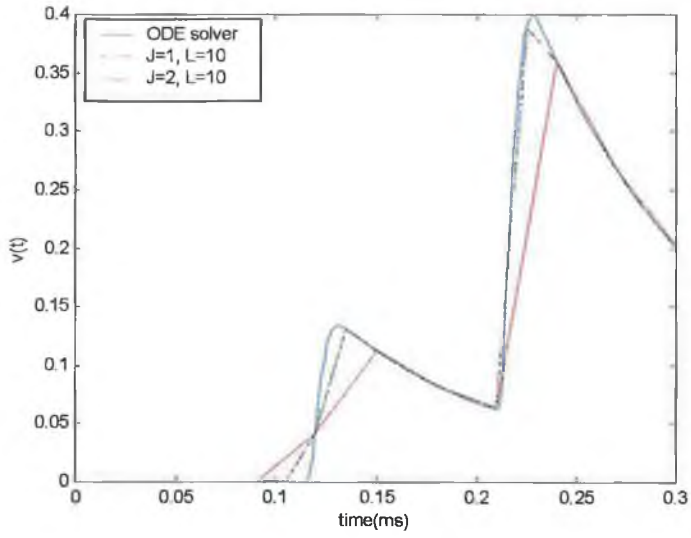


Fig. 9.1. *Accuracy improved by adding an extra layer (J=2) in wavelet series approximation*

The significant improvement in the accuracy of the circuit response, as evidenced from Fig. 9.1, confirms the rationale for employing extra layers. However, if the basic wavelet approach of Chapter 8 for simulating a system is employed, the

addition of extra layers increases the computational requirements greatly. But with the novel technique proposed in this Chapter, this is no longer a barrier.

Fig. 9.2 shows the results for the diode rectifier circuit with a new layer added $(J_1 = J + I = 2)$. The full line represents the result obtained using the full wavelet scheme with model reduction. The dashed line is the circuit response calculated at the same wavelet level but reusing results calculated from the lower-order simulation. As can be seen, these two responses are practically indistinguishable.

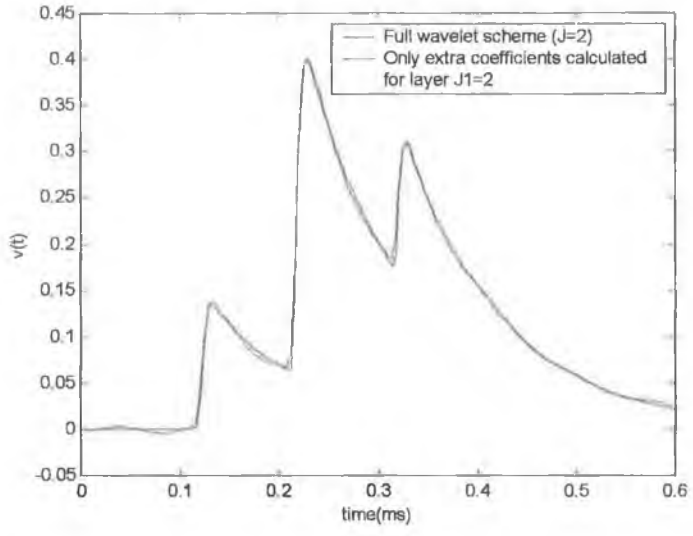


Fig. 9.2. *Result from the proposed new higer-order technique after adding an extra layer (J_1=2) in wavelet series approx.*

However, it took only 14% of the computing time to obtain the circuit response with a higher-degree of accuracy when compared to the time necessary to compute the circuit response by simply restarting the full wavelet simulation scheme with $J=2$.

9.2.2. MESFET amplifier

Fig. 9.3 presents the output obtained with the proposed new higher-degree accuracy technique after adding an extra layer ($J=2$) in the wavelet series approximation. It can be seen that the accuracy of the output voltage is considerably improved. However, it took only 21% of the computational time to obtain the circuit response with the new technique compared to the computational time required when the simulation is restarted from the beginning.

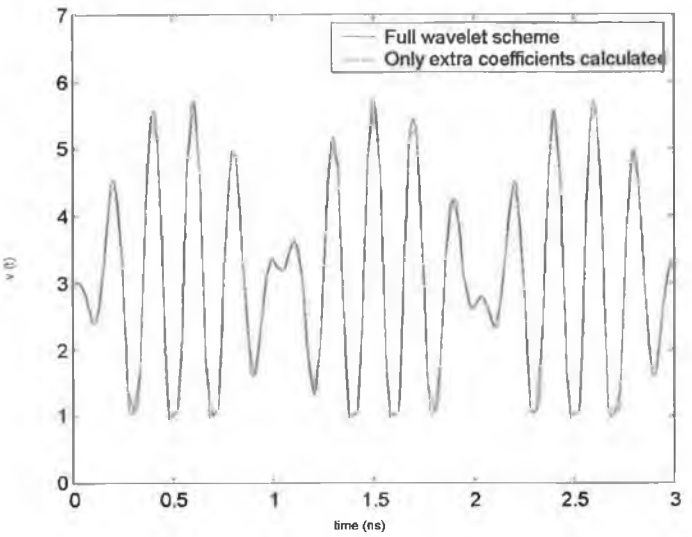


Fig. 9.3. *Result with the proposed new higer-order technique after adding an extra layer (J=2) in the wavelet series approximation*

Therefore, the results presented here clearly confirm that, by employing the approach presented here, the accuracy may be increased by adding an extra layer into the wavelet series approximation but with considerably less computational costs than restarting with a full wavelet scheme. This is possible since the coefficients calculated for a lower-order approximation are reused to form the higher-order approximation.

9.3. Further improvements for the IC design simulation technique

Equation (9.11) represents a $M \times M$ system of ODEs where the unknowns $\overline{\boldsymbol{g}}$ may be readily determined using any commercially available technique. However, as the degree of accuracy is increased by one layer, the number of additional coefficients M grows as a power of two. This in turn can drastically slow down the computation of the circuit response with higher-order accuracy. Therefore, it is desirable to reduce the size of this $M \times M$ system of ODEs before solving it.

Consider equations (8.9) and (9.11) that need to be solved in order to obtain the coefficients for the wavelet series expansion. As can be seen, the structure of these equations is exactly the same, only the entries in the corresponding matrices are different. Therefore, the same model order reduction technique as presented in Section 8.5.2. may readily be applied to the system in (9.11) yielding a new reduced equation system:

$$\tilde{E}_1 \frac{d\tilde{g}}{dt_1} = -\tilde{D}_1 \tilde{g} + Q_1^T F_M(Q_1 \tilde{g}) + \tilde{B}_M \tag{9.14}$$

where

$$\tilde{E}_1 = Q_1^T E_1 Q_1, \tilde{D}_1 = Q_1^T D_1 Q_1 \text{ and } \tilde{B}_M = Q_1^T B_M. \tag{9.15}$$

Again, the matrix Q_1 is obtained from orthogonal decomposition of a Krylov subspace formed from the coefficients of an Taylor series expansion of the vector of coefficients, $\hat{g}(t_1)$.

Thus instead of solving an M^{th} order system at each time-step to obtain the unknown state-variables, a reduced-order system of transformed coefficients is solved. The order of the reduced system q_1 is significantly less than M. Once the transformed coefficients are determined for the entire time range of interest, the additional M coefficients, $\hat{g}(t_1)$ and consequently, the value of the state variables and the output quantity $x(t)$ may be obtained in one single post-processing step. As a result, even more gains in computational efficiency are achieved as is confirmed for sample diode rectifier circuit given in Fig. 8.2.

9.3.1. Numerical results for a sample system

Fig. 9.4 presents the output of the sample diode rectifier circuit given in Fig. 8.2. The solid line is the circuit response when no MOR technique is applied to calculate the coefficients from added layer ($J_1=2$). The collocation points range parameter, L, is set to $L=80$.

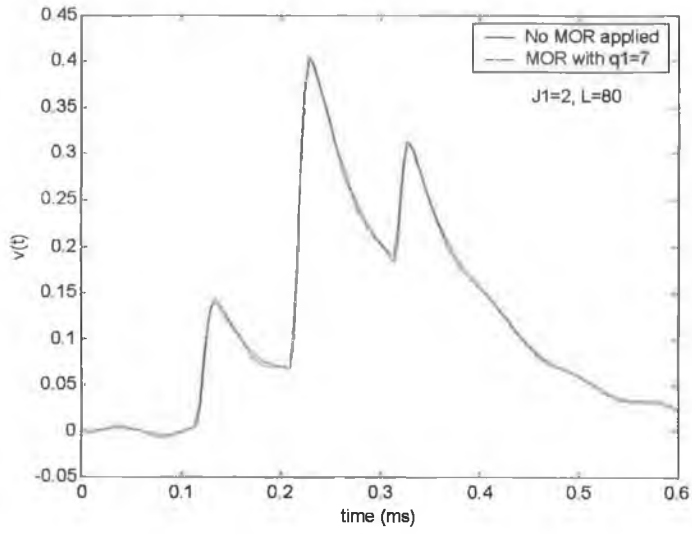


Fig. 9.4. *Result from proposed new technique with MOR applied*

As reported in Section 9.2.1, it took only 14% of the computing time to obtain the total higher degree accuracy circuit response when compared to the time necessary to compute the total circuit response by simply restarting the full wavelet simulation scheme at the same order of accuracy (*J=2).*

The dashed line in Fig 9.4 shows an output of the diode rectifier circuit using the enhanced technique proposed in this section. Parameters $J_1 = 2$ and $L = 80$ are the same as before and the system (9.11) is reduced to $q_1 = 7$. As reported in [DCB04b], it took only 9% of overall computing time to obtain the complete solution, which represents an additional efficiency improvement of 5%. This additional gain in computational efficiency is due to the fact that reduced system (9.14) with only 7 unknowns is solved using a standard ODE solver and the values for all coefficients in the extra layer are obtained in a single post-processing step involving only matrix multiplication.

9.4. Conclusion

Utilizing the multiresolution nature of wavelets, this chapter presents a further step towards a more accurate simulation technique with the potential to greatly shorten the IC design cycle. Rather than recalculating a complete set of new coefficients for a higher-order approximation of the unknown variable in the multi-time partial differential equation representation of the system, it utilises the coefficients calculated from a previous simulation that involved a lower-order approximation. Therefore, the technique can be very useful for the IC designer since it enables a desired accuracy requirement to be achieved in steps rather than restarting simulations each time a higher degree of accuracy is sought. Finally, the efficiency of this method is further improved by also using a non-linear model order reduction technique in the process for obtaining the wavelet coefficients for the extra layer in a higher-degree approximation.

CHAPTER 10

Conclusions

The aim of the research presented in this dissertation is to *advance the state-of-art in transient simulation of complex electronic circuits and systems operating at ultra high frequencies.* Highly accurate and efficient techniques for the simulation of linear interconnect networks with frequency-dependant parameters have been presented in the first part of the thesis. A novel wavelet-based strategy for the simulation of non-linear circuits subject to RF modulated signals has been developed and presented in the second part. Illustrative examples for both linear interconnects and non-linear circuits are presented to confirm both the efficacy and accuracy of the proposed strategies.

Chapter 1 introduces the research area. A comprehensive, but by no means exhaustive, list of the most important *challenges* facing the EDA community are summarised. The two main categories of *commercially available simulators,* circuit and full-wave simulators, are mentioned and it is underlined that the research efforts presented here are concerned with circuit simulators. The *main research objective* is stated: determining the transient response of a high-frequency complex system consisting of a linear and nonlinear part with greatly ***improved computational efficiency*** and with ***high accuracy.*** The approach should also permit an effective ***trade-off*** between accuracy and computational complexity.

Some important issues in relation to the design and simulation of high-speed circuits are reviewed in Chapter 2. *High-speed interconnect effects* such as propagation delay, rise-time degradation, attenuation, reflection and ringing, crosstalk and current distribution related effects are described. Their influence on the degradation of a propagating signal and the need for taking them into account in the early stages of circuit design is clearly illustrated. A short review of *existing interconnect models* (lumped, distributed transmission-line models, models based on tabulated data and full-wave models) is given and their merits and demerits are stated. Finally, important *simulation issues* relating to interconnect networks are stated.

Existing techniques for interconnect simulation are studied in Chapter 3. They may be classified as follows: *strategies based on transmission-line macromodelling* (lumped segmentation technique, direct time-stepping scheme, convolution techniques, the method of characteristics, exponential matrix rational approximation, basis function approximation, compact-finite-differences approximation and integrated congruence transform) and *model order reduction* based techniques (explicit moment-matching techniques such as Asymptotic Waveform Evaluation and Complex Frequency Hopping and Krylov subspace techniques such as the Arnoldi and Lanczos processes). In relation to MOR techniques, the important issues of stability, ill-conditioning of large matrices and passivity are briefly described.

In Chapter 4, a detailed *description of the resonant model* is given. The model is capable of providing an accurate description of a non-uniform line in the frequency-domain where the frequency-dependant parameters can be taken into account. The following particular advantages of the resonant model are identified: 1) an accurate frequency-domain prototype converts to a time-domain counterpart with minimal loss of accuracy and without the need for numerical convolution 2) the bandwidth of the model is explicit, i.e. the frequency components are accurately modelled up to a certain predetermined frequency 3) the particular structure of the model is such that it facilitates application of a model order reduction (MOR) algorithm thus improving the efficiency of the numerical calculations.

Two *novel model order reduction based techniques for interconnect modelling* are developed. The *first technique* is based on neglecting the higher order modes of propagation on the transmission line. This technique is straightforward to implement and excellent accuracy is retained even with more than a 50% reduction in the size of the original model. The number of modes to be neglected is determined by highest frequency that is required to be represented.

The *second technique, Lanczos MOR-based,* is developed to overcome the issues related to the high overall order of the $\mathcal{Z}$-domain admittance description that results from the resonant model. As evidenced by results, the technique is both accurate and numerically efficient. The method suffers from a common drawback of all Krylov-subspace techniques – determination of the reduction level is not an automated task. Although research efforts into overcoming this drawback are continuing at the moment, there is no solution for this problem that can be practically implemented.

The complexity and inhomogeneity of modern interconnect geometries is such that an analytical model cannot be formed and an alternative approach to the simulation of such interconnect is needed. To this end, a *novel technique* for the simulation of interconnects *described by set of tabulated data* is proposed in Chapter 5. The first stage involves enforcing causality by employing Hilbert Transform relationships. An FIR filter representation of network parameters is then synthesised and the Laguerre model order reduction process is employed in order to ensure the numerical efficiency of the new method. Experimental results that confirm both the accuracy and efficiency of the proposed approach are given. Since the technique is based on a set of tabulated data, it may be employed for large complex and/or inhomogeneous interconnect structures for which an analytical model would be too complicated or impossible to obtain.

Four new numerical algorithms for the transient analysis of high frequency nonlinear circuits are presented in Chapter 6. The algorithms address the issue of obtaining a solution to the *'stiff' ordinary differential equations* that arise in RF systems. The presented singlestep methods (Padé-Taylor and Padé-Xin) require obtaining analytical expressions for the higher order derivatives of the function governing the system. On the other hand, the multistep methods (Exact fit and Padé fit) introduced in this thesis do not require obtaining higher order derivatives but necessitate the use of a singlestep method to calculate values for the first few time-steps. Their use is recommended in cases of very complicated analytical functions. Finally, corrector formulas for use in predictor-corrector schemes are also proposed.

Chapter 7 presents an introduction to wavelets and wavelet theory as well as the rationale for the use of wavelet functions as a basis for developing a novel envelope transient simulation technique. The relationship between the Wavelet transform (WT) an the Fourier transform (FT) is highlighted and some essential wavelet properties are presented. The discrete wavelet transform (DWT) is suggested for the purpose of efficient numerical implementation of the Wavelet Transform. Finally, the detailed definition of scaling and wavelet functions as well as some basic properties of a *wavelet-like multiresolution collocation scheme* are presented.

This multiresolution collocation scheme forms a core of a novel wavelet-based technique for envelope transient simulation that is described in Chapter 8. The technique utilizes the multi-time partial differential approach in combination with

wavelet-basis functions. *A non-linear model reduction scheme* is also employed resulting in significant gains in terms of computational efficiency. A particular advantage of the proposed technique is that it enables a simple trade-off between the required accuracy and the desired efficiency of the computational algorithm. Since the proposed wavelet basis exhibits good approximation properties to the unknown variables, very good accuracy may be achieved employing only a shallow wavelet level. Selection of the wavelet level requires care as use of a deep level where unnecessary results in ill-conditioning.

An efficient nonlinear circuit simulation technique *with the potential to significantly reduce the overall design cycle* is presented in the Chapter 9. The key factor is the structure of the wavelet-based technique presented in Chapter 8. It enables reuse of the previously calculated transient response results to calculate a more accurate response but without the need to restart the simulation from the beginning. This is a particularly useful feature, e.g. when fine-tuning of an initial design is required.

To conclude, this thesis has addressed the issue of obtaining ***highly accurate*** transient responses of a high-frequency complex system consisting of linear and non-linear parts with greatly ***improved computational efficiency*** in a way that permits an effective ***trade-off*** between accuracy and computational complexity.

Several issues have been identified as areas for possible extensions to the research presented in this dissertation. These include: the choice of the most suitable linear model order reduction technique, the choice of the optimal wavelet basis set, the implementation of proposed ODE solvers for large stiff systems and the coding of the proposed methods in a compiler based language (e.g. C++). These and some of the related research areas are discussed in the remainder of this chapter.

A Lanczos-based linear model reduction technique is used to improve the numerical efficiency of the analytical resonant model (Chapter 4) while a Laguerre based linear MOR technique is incorporated into the simulation technique for interconnects described by a tabulated set of data (Chapter 5). Although the chosen MOR techniques give good results, it is not proven that they are the optimal ones. Therefore, a further investigation into the available linear model order techniques with a

view to identifying the optimal MOR scheme for the proposed interconnect simulation technique is suggested.

A wavelet-like basis set for solving the initial boundary value problems as proposed by Cai and Wang [CW96] is used in this thesis. This particular wavelet-like system has been chosen because of its superior capabilities in dealing with strong non-linearities. However, there is a need to explore different wavelet bases to ascertain the most effective bases for use within the proposed envelope technique.

The methods for solving an ordinary differential equation proposed in this thesis involve using Padé approximants to achieve accuracy while speeding-up calculations. The speed-up is accomplished by enabling the use of a longer time-step when compared to the traditional ODE solvers. The methods are tested on a single ODE and on a simple system of ODEs and as observed, the initial results are encouraging. Application of the proposed methods to very large systems of ODEs as arise from mathematical models of industrial high-speed electronic circuits is necessary.

The techniques for simulation of a complex electronic circuit presented in this thesis are shown to be very effective for the simulation of small-scale electronic circuits. However, the techniques need to be tested on large-scale complex electronic circuits. For that purpose, the proposed algorithms that are coded in MATLAB language need to be implemented in a simulation platform that enables obtaining results in real time. Since MATLAB is an interpreter language, the algorithm execution time is much longer than if the same algorithm was coded in a compiler languages, e.g. C++. Therefore, the methods presented in this thesis need to be implemented in a simulation platform that enables the technique to be compared to existing techniques in terms of accuracy and efficiency and their trade-off.

Circuit simulation is almost as old as IC design. Both need to develop in parallel in order to ensure further progress in the field. Shrinking device sizes and the constant rise in the operational frequency of chips necessitate reliable and robust simulation algorithms. The major limiting factor in IC performance is the effect of the interconnect network and this factor is of paramount importance now with clock speeds well into the gigahertz frequency range and signals with picosecond rise times. Furthermore, interconnects now have such complex topologies and geometries that there is significant

coupling between many physical levels. Thus the issue of efficient and accurate inclusion of all interconnect effects at all levels of the design process is of great importance for developers of EDA tools.

Coupling between models and algorithms from different domains (e.g. linear and non-linear, analog and digital, thermal and electrical) is another question of great interest. Very often, time constants related to such domains vary greatly thus making computations decidedly inefficient. Thus, an accurate and efficient multi-rate solver is needed in order to yield simulation results in an acceptable amount of time.

In short, an efficient and accurate computer-aided design tool of the future has to be able to handle very complex non-linear circuits incorporating accurate models for large interconnect networks without putting too much stress on the CPU and memory requirements.

BIBLIOGRAPHY

[A04] G. Antonini, "New methodology for the transient analysis of lossy and dispersive multiconductor transmission lines", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 52, Sep. 2004, pp. 2227-2239.

[AN01] R. Achar and M.S. Nakhla, "Simulation of high-speed interconnects", *Proc. IEEE*, Vol. 89, May 2001, pp. 693-728.

[AN04] R. Achar and M. Nakhla, "Managing complexity of high-speed modules via model-order reduction", *Notes on TFB tutorial, International Microwave Symposium (IMS)*, Fort Worth, TX, USA, June 2004.

[ANZ98] R. Achar, M. Nakhla and Q.J. Zhang, "Full-wave analysis of high-speed interconnects using complex frequency hopping", *IEEE Trans. on Computer-Aided Design*, Vol. 17, Oct. 1998, pp. 997–1016.

[AP98] U.M. Ascher and L.R. Petzold, "*Computer methods for ordinary differential equations and differential-algebraic equations*", Philadelphia, PA: SIAM, 1998.

[AS00] A.C. Antoulas and D.C. Sorensen, "Approximation of large-scale dynamical systems", *SIAM Annual Meeting*, San Juan, Puerto Rico, July 2000.

[AW95] K. Amaratunga and J. R. Williams, "Time integration using wavelets," [Online], 1995, also in *Proc. SPIE - Wavelet applications for dual use*, Orlando, FL, USA, April 1995, pp. 894-902. [Accessed 20 December 2004].
Available: http://wavelets.mit.edu/Kevin/Papers/Waveextrap-ivp.pdf

[B00] J.C. Butcher, "Numerical methods for ordinary differential equations in the 20th century", *Journal of Computational and Applied Mathematics*, Vol. 125, Dec. 2000, pp. 1-29.

[B00] W. Bandurski, "Simulation of single and coupled transmission lines using time-domain scattering parameters", *IEEE Trans. on Circuits and Systems I*, Vol. 47, Aug. 2000, pp. 1224-1234.

[B67] F.H. Branin, "Transient analysis of lossless transmission line", *Proceedings of the IEEE,* Vol. 55, Nov 1967, pp. 2012-2013.

[B87] G. Battle, "A block spin construction of ondelettes. Part I: Lemarie functions.", *Communications in Mathematical Physics*, Vol. 110, 1987, 601-615.

[B90] H.B. Bakoglu, "*Circuits, interconnections and packaging for VLSI*", Reading, MA: Addison – Wesley, 1990.

[B94] D. L. Boley, "Krylov space methods on state-space control models," *Circuits Systems Signal Processing*, vol. 13, 1994, pp. 733–758.

[B95] T.J. Brazil, "Causal convolution – a new method for the transient analysis of linear systems at high frequencies", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 43, Feb 1995, pp. 315-323

[B98] B. Burke-Hubbard, "*The world according to wavelets: the story of the mathematical technique in making*", 2nd Edition, Natick, MA: AK Peters Ltd, 1998.

[BA883] F. Bashforth and J.C. Adams, "*An attempt to test the theories of capillary action by comparing the theoretical and measured forms of drops of fluid, with an explanation of the method of integration employed in constructing the tables which give the theoretical forms of such drops*", Cambridge: Cambridge University Press, 1883.

[BG81] G.A. Baker, Jr and P. Graves-Morris, "Padé approximants, Part I: Basic theory", Reading, MA: Addison – Wesley, 1981.

[BGG98] C.S. Burrus, R.A. Gopinath and H. Guo, "Introduction to wavelets and wavelet transforms: A primer", New Jersey: Prentice Hall, 1998.

[BKG00] N. Boulejfen, A. Kouki and F. Ghannouchi, "Frequency and time domain analysis nonuniform lossy coupled transmission lines withlinear and nonlinear terminations", *IEEE Trans. on Microwave Theory and Techniques,* Vol. 48, March 2000, pp. 367–379.

[BR00] S. Barmada and M. Raugi, "Transient numerical solutions of nonuniform MTL equations with nonlinear loads with wavelet expansion in time or space-domain," *IEEE Trans. on Circuits and Systems I*, Vol. 47, Aug. 2000, pp. 1178–1190.

[BS97] W.T. Beyene and J.E. Schutt-Aine, "Accurate frequency-domain modelling and efficient simulation of high-speed packaging interconnects", *IEEE Trans. on Microwave Theory and Techniques,* Vol. 45, Oct. 1997, pp. 1941-1947.

[BS98] W.T. Beyene and J.E. Schutt-Aine, "Efficient transient simulation of high-speed interconnects characterized by sampled data," *IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging*, [see also IEEE Trans. on Components, Hybrids, and Manufacturing Technology], Vol. 21, Feb. 1998, pp. 105–113.

[BSK01] V. Balakrishnan, Q. Su and C-K. Koh, "Efficient balance and truncate model reduction for large scale systems", in *Proc. IEEE American Control Conference*, Arlington, USA, June 2001, pp. 4746-4751.

[C01] M. Condon, "A new technique for the transient simulation of transmission lines inclusive of skin effect", in *Proc. European Microwave Conference*, London, Sep. 2001.

[C02a] M. Condon, "Time domain simulation of coupled interconnects", in *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering (COMPEL)*, Vol. 21, July 2002, pp. 438-450

[C02b] M. Condon, "Efficient transient simulation of transmission lines", in *IEEE MTT-S International Microwave Symposium Digest*, Vol. 3, Seattle, June 2002, pp. 2065-2068.

[C03] I. Christov, "Wavelet-Galerkin methods for partial differential equations", [Online], July 2003, presented at Massachusetts Institute of Technology Matrix Analysis and Wavelets REU, Texas A&M University Summer 2003, [Acessed 19 December 2004].
Available: http://stuff.mit.edu/people/christov/downloads/wg-pres.pdf

[C04] Chou S, "Extending Moore's law in the nanotechnology era" [Online], 2004, [Accessed 4 March 2004].
Available: http://www.intel.com/research/silicon/nanotechnology.htm

[C89] F.-Y. Chang, "The generalized method of characteristics for waveform relaxation analysis of lossy coupled transmission lines", *IEEE Trans. on Microwave Theory and Techniques*, vol. 37, Dec. 1989, pp. 2028–2038.

[C92] R.E. Collin, "*Foundation for microwave engineering*", 2nd Edition, Wiley-IEEE Press, 2000.

[C98] M. Condon, "*Application of resonant analysis to the modelling of transformers and transmission lines*", Ph.D. dissertation, The Department of Electronic Engineering, National University of Ireland, Galway, 1998.

[CBK+02] Y. Chen, V. Balakrishnan, C-K. Koh and K. Roy, "Model reduction in the time-domain using Laguerre polynomials and Krylov methods", in *Proc. Design Automation and Test in Europe (DATE)*, Paris, France, Mar. 2002, pp. 931-935.

[CBM+02] K. M. Coperich-Branch, J. Morsey, A.C. Cangellaris, and A.E. Ruehli, "Physically consistent transmission line models for high-speed interconnects in lossy dielectrics", *IEEE Trans. on Advanced Packaging*, Vol. 25, May 2002, pp. 129-135.

[CC97] M. Celik and A.C. Cangellaris, "An all purpose transmission line model for interconnect simulation in SPICE," *IEEE Trans. on Microwave Theory and Techniques*, Vol. 45, Oct 1997, pp. 1857-1867.

[CC98] A.C. Cangellaris and M. Celik, "Order reduction of high-speed interconnect electrical models: The issue of passivity", in *Proc. IEEE Symposium on IC/Package Design Integration*, Santa Cruz, CA, Feb. 1998, pp. 132–137.

[CCH+01] A.C. Cangellaris, K.M. Coperich, H-M. Huang and A. Ruehli, "Investigation of the numerical accuracy of transmission-line based transient analysis of high-speed interconnects and cables", in *Proc. IEEE International Symposium on Electromagnetic Compatibility*, Montreal, Canada, Vol. 2, Aug. 2001, pp. 821– 826.

[CD03] M. Condon and E. Dautbegovic, "A novel envelope simulation technique for high-frequency nonlinear circuits", in *Proc. 33rd European Microwave Conference (EuMC)*, Munich, Germany, Vol. 2, October 2003, pp. 619-622.

[CD03a] M. Condon and E. Dautbegovic, "Efficient modeling of interconnects in high-speed circuits", in *Proc. European Conference on Circuit Theory and Design (ECCTD)*, Vol. 3, Kraków, Poland, September 2003, pp. 25-28.

[CDB02] M. Condon, E. Dautbegovic and T.J. Brazil, "An efficient numerical algorithm for the transient analysis of high frequency non-linear circuits", in *Proc. 32nd European Microwave Conference (EuMC)*, Milan, Italy, Sep. 2002.

[CDB05] M. Condon, E. Dautbegovic and C. Brennan, "Efficient transient simulation of interconnect networks", to be published in *COMPEL journal*, 2005.

[CLC00] C.K. Cheng, J. Lillis, S. Lin and N. Chang, "*Interconnect Analysis and Synthesis*", New York: John Wiley and Sons, 2000.

[CM02] S. M. Cox and P. C. Matthews, "Exponential time differencing for stiff systems", *Journal of Computational Physics*, Vol. 176, March 2002, pp. 430-455.

[CN94] E. Chiprout and M.S. Nakhla, "*Asymptotic waveform evaluation and moment matching for interconnect analysis*", Boston: Kluwer Academic Publishers, 1994.

[CN95] E. Chiprout and M.S. Nakhla, "Analysis of interconnect networks using complex frequency hopping (CFH)", *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 14, Feb. 1995, pp. 186-200.

[CPO02] M. Celik, L. Pileggi and A. Odabasioglu, "*IC Interconnect Analysis*", Boston: Kluwer Academic Publishers, 2002.

[CPP+99] C. Cangellaris, S. Pasha, J. L. Prince and M. Celik, "A new discrete time-domain model for passive model order reduction and macromodeling of high-speed interconnections," *IEEE Trans. on Advanced Packaging* [see also IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging], Vol. 22, Aug. 1999, pp. 356–364.

[CRZ00] J. Cullum, A. Ruehli, and T. Zhang, "A method of reduced-order modeling and simulation of large interconnect circuits and its application to PEEC models with retardation," *IEEE Trans. on Circuits and Systems,* Vol. 47, Apr. 2000, pp. 261–273.

[CS01] C.E. Christoffersen and M.B. Steer, "State-variable-based transient circuit simulation using wavelets", *IEEE Microwave and Wireless Components Letters*, Vol. 11, April 2001, pp. 161-163

[CW00] Y. Chen and J. White, "A quadratic method for nonlinear model order reduction", in *Proc. International Conference on Modelling and Simulation of Microsystems Semiconductors, Sensors and Actuators*, San Diego, USA, March 2000, pp. 477-480.

[CW92] R.R. Coifman and M.V. Wickerhauser, "Entropy based algorithms for best basis selection", *IEEE Trans. on Information Theory*, Vol. 38, Mar 1992, pp. 713-718.

[CW96] W. Cai and J.Z. Wang, "Adaptive multi-resolution collocation methods for initial boundary value problems of nonlinear PDEs", *SIAM Journal Numerical Analysis*, Vol. 33, June 1996, pp. 937-970

[D04] G. Denk, "Circuit simulation for nanoelectronics", presented at *SCEE 2004*, Capo D'Orlando, Italy, Sep. 2004.

[D88] Daubechies, "Orthonormal bases of compactly supported wavelets", *Communications on Pure and Applied Mathematics*, vol. 41, Nov. 1988, pp. 909-996.

[D92] Daubechies, "*Ten lectures on wavelets*", CBMS-NSF Regional Conference Series in Applied Mathematics, No. 61, Philadelphia: SIAM 1992.

[D93] D.L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation", *Applied and computational harmonic analysis*, Vol. 1, Dec 1993, pp. 100-115.

[D97] J. W. Demmel, "*Applied Numerical Linear Algebra*", Philadelphia, PA: SIAM, 1997.

[D98] A. Deutsch, "Electrical characteristics of interconnections for high performance systems," *Proc. IEEE*, Vol. 86, Feb. 1998, pp. 315–355.

[D99] W. Dahmen, "Wavelet methods for PDEs - some recent developments", [Online], Dec. 1999, *IGPM Report No. 183*, RWTH Aachen, [Accessed 11 November 2004].
Available: ftp://ftp.igpm.rwth-aachen.de/pub/dahmen/2000_5.ps.Z

[DAR02] "*EDA design automation roadmap*", [Online], Executive summary, 3rd release, 2002, [Accessed 24 Nov 2004].
Available: http://www.medea.org/webpublic/pdf/EDA_ExecSum_030310.pdf

[DAR03a] "*Design automation roadmap*", [Online], Executive summary, 4th release, 2002, [Accessed 24 November 2004].
Available: http://www.medea.org/webpublic/pdf/2003RoadmapExec.pdf

[DAR03b] "*Design automation solutions for Europe*", [Online], 3rd release, 2002, [Accessed 24 November 2004].
Available: http://www.medea.org/webpublic/pdf/EDA_roadmap_2002.pdf

[DB78] C.R. DeBoor, "*A practical guide to splines*", New York: Springer-Verlag, 1978.

[DC03] E. Dautbegovic and M. Condon, "Efficient simulation of interconnects in high-speed circuits", in *Proc. 8th IEEE High-Frequency Postgraduate Colloquium*, Belfast, UK, September 2003, pp. 81-84.

[DCB04a] E. Dautbegovic, M. Condon and C. Brennan, "An efficient nonlinear circuit simulation technique", in *Proc. IEEE Radio Frequency Integrated Circuit Symposium (RFIC)*, Fort Worth, TX, USA, June 2004, pp. 623-626.

[DCB04b] E. Dautbegovic, M. Condon and C. Brennan, "An efficient wavelet-based nonlinear circuit simulation technique with model order reduction", in *Proc. 9th IEEE High-Frequency Postgraduate Colloquium*, Manchester, UK, September 2004, pp. 119-124.

[DCB05] E. Dautbegovic, M. Condon and C. Brennan, "An efficient nonlinear circuit simulation technique", in *IEEE Transactions on Microwave Theory and Techniques*, February 2005, pp.548 – 555.

[DCK+01] A. Deutsch, P.W. Coteus, G.V. Kopcsay, H.H. Smith, C.W. Surovic, B.L. Krauter, D.C. Edelstein and P.L. Restle, "On-chip wiring design challenges for gigahertz operation", *Proc. IEEE*, Vol. 89, Apr. 2001, pp. 529 – 555.

[DKC+01] A. Deutsch, G.V. Kopcsay, P.W. Coteus, C.W. Surovic, P.E. Dahlen, D.L. Heckmann and D-W. Duan, "Frequency-dependent losses on high-performance interconnections", *IEEE Trans. on Electromagnetic Compatibility*, Vol. 43, Nov. 2001, pp. 446 – 465.

[DKO97] W. Dahmen, A. Kurdila and P. Oswald, "Multiscale wavelet for partial differential equations", in *Wavelet analysis and its applications,* vol. 6, Ed. Charles K. Chui, San Diego: Academic Press, 1997.

[DKR+97] Deutsch, G.V. Kopcsay, P.J. Restle, H.H. Smith, G. Katopis, W.D. Becker, P.W. Coteus, C.W. Surovic, B.J. Rubin, R.P. Dunne, Jr., T. Gallo, K.A. Jenkins, L.M. Terman, R.H. Dennard, G.A. Sai-Halasz, B.L. Krauter and D.R. Knebel, "When are transmission-line effects important for on-chip interconnections?", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 45 , Oct. 1997, pp. 1836 – 1846.

[DMD92] T. Dhaene, L. Martens and D. DeZutter "Transient simulation of arbitrary nonuniform interconnection structures characterised by scattering parameters", *IEEE Trans. on Circuit and Systems -I: Fundamental Theory and Applications*, Vol. 39, Nov 1992, pp. 928-937.

[DNA00] R. Dounavis, R. Achar and M. Nakhla, "Efficient passive circuit models for distributed networks with frequency-dependent parameters," *IEEE Trans. on Advanced Packaging* [see also IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging], vol. 23, Aug. 2000, pp. 382–392.

[DNA99] R. Dounavis, X. Li, M. Nakhla, and R. Achar, "Passive closed-loop transmission line model for general purpose circuit simulators," *IEEE Trans. on Microwave Theory and Techniques*, vol. 47, Dec. 1999, pp. 2450–2459.

[DS87] A.R. Djordjević and T.K. Sarkar, "Analysis of time response of lossy multiconductor transmission line networks", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 35, Oct. 1987, pp. 898 – 908.

[DS94] A. R. Djordjević and T. K. Sarkar, "Closed-form formulas for frequency-dependent resistance and inductance per unit length of microstrip and strip transmission lines," *IEEE Trans. on Microwave Theory and Techniques*, Vol. 42, Feb. 1994, pp. 241–248.

[DSB86] A. R. Djordjević, T. K. Sarkar and R.F. Barrington, "Analysis of lossy transmission lines with arbitrary nonlinear terminal networks", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 34, Jun. 1986, pp. 660 – 666.

[DSS+99] A. Deutsch, H.H. Smith, C.W. Surovic, G.V. Kopcsay, D.A. Webber, P.W. Coteus, G.A. Katopis, W.D. Becker, A.H. Dansky, G.A. Sai-Halasz and P.J. Restle, "Frequency-dependent crosstalk simulation for on-chip interconnections", *IEEE Trans. on Advanced Packaging* [see also IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging], Vol. 22, Aug. 1999, pp. 292–308.

[DZ92] T. Dhaene and D.D. Zutter, "Selection of lumped element models for coupled lossy transmission lines," *IEEE Trans. on Computer-Aided Design*, vol. 11, July 1992, pp. 805-815.

[E04] C. Evans-Pughe, "Synthesis and the new physics", *IEE Review*, Special Report on Electronic Design, Aug. 2004, pp.28-31.

[E48] W. C. Elmore, "The transient response of damped linear networks with particular regard to wide-band amplifiers," *Journal of Applied Physics*, Vol. 19, Jan. 1948, pp. 55–63.

[EB03] M.A. Elgamel and M.A. Bayoumi, "Interconnect noise analysis and optimization in Deep submicron technology", *IEEE Circuits and Systems Magazine*, Vol. 3, 2003, pp. 6-17.

[EHR+02] M.Elfadel, H.M. Huang, A.E. Ruehli, A. Dounavis and M.S. Nakhla, "A comparative study of two transient analysis algorithms for lossy transmission lines with frequency-dependent data", *IEEE Trans. on Advanced Packaging* [see also IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging], Vol. 25, May 2002, pp. 143 – 153.

[EL97] I.M. Elfadel and D.D. Ling, "A block rational Arnoldi algorithm for multi-port passive model-order reduction for multiport RLC networks", in *Digest IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, Nov. 1997, pp. 66-71.

[F93] J. B. Faria, "*Multiconductor transmission line structures*", New York: Wiley, 1993.

[FF95a] P. Feldmann and R. W. Freund, "Efficient linear circuits analysis by Pade approximation via the Lanczos process", *IEEE Trans. on Computer-Aided Design*, Vol. 14, May 1995, pp. 639-649.

[FF95b] P. Feldmann and R. W. Freund, "Reduced-order modelling of large linear subcircuit via a block Lanczos algorithm", in *Proc. IEEE/ACM Conference on Design Automation (DAC)*, San Francisco, CA, June 1995, pp. 474–479.

[FF96] R.W.Freund and P.Feldmann, " Reduced-order modeling of large passive linear circuits by means of the SyPVL algorithm", *Proc. IEEE/ACM Conference on Design Automation (DAC)*, Las Vegas, Nevada, June 1996, pp. 280-286.

[G71] C.W. Gear, "*Numerical initial value problems in ordinary differential equations*", New Jersey: Prentice-Hall, 1971.

[G79] A. Graham, "*Matrix theory and applications for engineers and mathematicians*", Chichester: Ellis Horwood limited, 1979.

[G84] K. Glover, "Optimal Hankel-norm approximations of a linear multivariable systems and their L^{∞} -error bounds", *International Journal of Control*, Vol. 39, 1984, pp. 1115-1193.

[G94] R. Goyal, "Managing signal integrity," *IEEE Spectrum*, Mar. 1994, pp. 54–62.

[GA00] S. Gugercin and A. C. Antoulas, "A comparative study of 7 model reduction algorithms", *Proc. 39th IEEE Conference on Decision and Control*, Vol. 3, Sydney, Australia, Dec. 2000, pp. 2367 – 2372.

[GC01] S. Grivet-Talocia and F. Canavero, "Wavelet-based high-order adaptive modelling of lossy interconnects", *IEEE Trans. on Electromagnetic Compatibility*, Vol. 43, Nov. 2001, pp. 471–484.

[GLR+90] R. Glowinski, W. Lawton, M. Ravachol and E. Tenenbaum, "Wavelet solutions of linear and non-linear elliptic, parabolic, and hyperbolic problems in one space dimension", *Computational Methods in Applied Sciences and Engineering*, Ed. R. Glowinski, A. Lichnecosky, SIAM, 1990.

[GM04] E. Gad and M. Nakhla, "Efficient simulation of nonuniform transmission lines using integrated congurence transform", *IEEE Trans. on VLSI systems*, Vol. 12, Dec. 2004, pp. 1307-1320.

[GM84] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", *SIAM Journal of Mathematical Analysis*, Vol. 15, July 1984, 723-736.

[GN01] P. Gunupudi and M. Nakhla, "Non-linear circuit reduction of high-speed interconnect networks using congruent transformation techniques", *IEEE Trans. on Advanced Packaging* [see also IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging], Vol. 24, Aug. 2001, pp. 317-325.

[GN97] R. Griffith and M. Nakhla, "A new high-order absolutely-stable explicit numerical integration algorithm for the time-domain simulation of non-linear circuits", *Digest of IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, Nov. 1997, pp. 276–280.

[GN99] P.K. Gunupudi and M.S. Nakhla, "Model reduction of nonlinear circuits using Krylov-subspace techniques", in *Proc. IEEE Design Automation Conference*, New Orleans, June 1999, pp. 13-16.

[GW99] C.F. Gerald and P.O. Wheatley, "*Applied numerical analysis*", Reading: Addison Wesley, 1999.

[H01] P. Heres, "*Reduced order modelling of RLC-networks using an SVD-Laguerre method*" Master's dissertation, Eindhoven University of Technology, 2001.

[H10] A. Haar "Zur Theorie der orthogonalen Funktionensysteme", *Mathematische Annalen*, vol. 69, 1910, pp. 331-371.

[H62] P. Henrici, "*Discrete variable methods in ordinary differential equations*", John Wiley & Sons, 1962.

[HKG+03] S. Hamid, T. Kolev, Q.T.L. Gia and W. Wu, "Numerical solution of partial differential equations using wavelet approximation space", [Online], May, 2003, MATH 667 project, [Accessed 11 December 2004].
Available: http://www.math.tamu.edu/~qlegia/Math/WAVE/wav_pde.ps.gz

[HNW87] E. Harier, S.P. Nørsett and G. Wanner, "*Solving ordinary differential equations I*", Springer-Verlag, 1987.

[HO04] M. Hochbruck and A. Ostermann, "Exponential Runge-Kutta methods for parabolic problems", [Online], preprint 2004, [Accessed 15 September 2004].
Available: http://techmath.uibk.ac.at/numbau/alex/publications/04-1.pdf

[HR91] H. Heeb and A. Ruehli, "Retarded models for PC board interconnects – or how the speed of light affects your SPICE simulation", in *Digest IEEE International Conference on Computer-Aided Design (ICCAD),* Santa Clara, CA, Nov. 1991, pp. 70–73.

[HRB75] C.W. Ho, A.E. Ruehli and P.A. Brennan, "The modified nodal approach to network analysis", *IEEE Trans. on Circuits and Systems*, Vol. 22, June 1975, pp.504-509.

[HS96] D.G. Haigh and R.S. Soin, "*High frequency circuit engineering*", IEE Circuits and Systems Series 6, 1996 (originally published in German by Expert-Verlag 1990).

[I04] *Icknowledge* [Online], 2004, [Accessed 10 October 2004].
Available: http://www.icknowledge.com/glossary/i.html

[IJ02] E.C. Ifeachor and B.W. Jervis, "*Digital Signal Processing: a practical approach*", 2nd edition, New York : Prentice Hall, 2002.

[ITRS99a] "*THE INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS: 1999 – Interconnects*", [Online], 1999, [Accessed 24 November 2004].
Available: http://public.itrs.net/files/1999_SIA_roadmap/ Int.pdf

[ITRS99b] "*THE INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS: 1999 – Modelling and simulation*", [Online], 1999, [Accessed 24 November 2004].
Available: http://public.itrs.net/files/1999_SIA_roadmap/Modeling.pdf

[J84] M.K. Jain, "*Numerical Solution of Differential Equations*", 2nd Edition, New Delhi: Wiley Eastern Limited, 1984.

[JG93] H. Johnson and M. Graham, "*High-speed Digital Design: A Handbook of Black Magic*". New York: Prentice Hall, 1993.

[K04a] A. Kvaerno, "Time integration method for coupled equations", [Online], 2004, Preprint Numerics No. 7/2004, [Accessed 07 January 2005].
Available: http://www.math.ntnu.no/preprint/numerics/2004/N7-2004.pdf

[K04b] S. Krogstad, "Generalized integrating factor methods for stiff PDEs", [Online], preprint 2004, [Accessed 15 September 2004].
Available: www.ii.uib.no/~stein/thesis/gif_t.pdf

[K63] R.E. Kalman, "Lyapunov functions for the problem of Lur'e in Automatic Control", in *Proc. National Academy of Science,* Vol. 256, 1963, pp. 201-205.

[K94] G. Kaiser, "*A Friendly Guide to Wavelets*", Boston: Birkhäuser, 1994.

[KGP94] S. Y. Kim, N. Gopal, and L. T. Pillage, "Time-domain macromodels for VLSI interconnect analysis," *IEEE Trans. on Computer-Aided Design*, vol. 13, Oct. 1994, pp. 1257–1270.

[KMS+98] M. Kamon, N.Marques, L.M Silveira and J.White, "Automatic generation of accurate circuit models of 3-D interconnect", *IEEE Trans on Components, Packaging and Manufacturing Technology – Part B*, Vol.21, Aug.1998, pp.225-240.

[KS86] K.S. Kundert and A. Sangiovanni-Vincentelli, "Simulation of nonlinear circuits in the frequency domain", *IEEE Trans. on Computer-Aided Design*, Vol.5, Oct. 1986, pp. 521-535.

[KT04] A.-K. Kassam and L. N. Trefethen, "Fourth-order time stepping for stiff PDEs", [Online], preprint 2004, [Accessed 20 December 2004].
Available: http://www.math.auckland.ac.nz/~will/papers/Kassam-Trefethen04.pdf

[KWY95] K. J. Kerns, I. L. Wemple and A. T. Yang, "Stable and efficient reduction of substrate model networks using congruence transforms," *Digest IEEE International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, Nov. 1995, pp. 207–214.

[KY97] K.J. Kerns and A. T. Yang, "Preservation of passivity during RLC network via split congruent transformations", in *Proc. IEEE/ACM Conference on Design Automation (DAC)*, Anaheim, CA, June 1997, pp. 34-39.

[KZ99] L. Knockaert and D. DeZutter, "Passive reduced order multiport modelling: The Padé-Arnoldi-SVD connection", in *International Journal of Electronics and Communications*, Vol. 53, 1999, pp. 254-260.

[L04] D. Lammers, "Looking past planar CMOS", [Online], Sep. 2004, *EE Times*, [Accessed 24 November 2004].
Available: http://www.eetimes.com/showArticle.jhtml?articleID=46800129

[L88] P. G. Lemarie, "Une nouvelle base d'ondettes de $L^2(R)$ ", J. *de Math. Pures et Appl.*, Vol. 67, 1988, pp. 227-236.

[LA89] Y. Liu and B. D. O. Anderson, "Singular perturbation approximation of balanced systems", *International Journal of Control*, Vol. 50, 1989, pp. 1379-1405.

[LK92] S.Lin and E.S. Kuh, "Transient simulation of lossy interconnects based on the recursive convolution formulation", *IEEE Trans. on Circuits and Systems I,* Vol. 39, Nov.1992, pp.879-892.

[LKM93] S.Lin, E.S Kuh and M.Marek-Sadowska, "Stepwise equivalent conductance circuit simulation technique", *IEEE Trans. on Computer-Aided Design*, Vol.12, May 1993, pp. 672-683.

[LKS93] S.Y. Lee, A. Konrad and R. Saldanha, "Lossy transmission line transient analysis by the finite element method" *IEEE Trans. on Magnetics*, Vol. 29, Mar. 1993, pp. 1730 – 1732.

[LMA+97] D. Long, R. Melville, K. Ashby and B. Horton, "Full-chip harmonic balance", *Proc. IEEE Custom Integrated Circuits Conference*, Santa Clara, CA, May 1997, pp. 379-382.

[LMG02] S. Lall, J.E. Marsden and S. Glavaski, "A subspace approach to balanced truncation for model reduction of nonlinear control sytems", *International Journal of Robust and Nonlinear Control*, Vol. 12, 2002, pp. 519-35.

[LRT91] A. Latto, H.L. Resnikoff and E. Tenenbaum, "The Evaluation of Connection Coefficients of Compactly Support Wavelets", [Online], 1991, also in Proc. Princeton Conference on Wavelets and Turbulence, [Accessed 19 December 2004].
Available: http://www2.appmath.com:8080/site/ftp/con3_5.pdf

[LT90] A. Latto and E. Tenenbaum, "Compactly supported wavelets and the numerical solution of Burgers' equation", [Online], 1990, also in *Comptes Rendus,* French Academy of Science, t. 311, Serie I, 1990, pp. 903-909, [Accessed 19 December 2004].
Available: http://www2.appmath.com:8080/site/ftp/burger.pdf

[M65] G.E. Moore, "Cramming More Components Onto Integrated Circuits", *Electronics*, Vol. 38, Apr. 1965, pp. 114-117.

[M69] R.E. Matick, "*Transmission lines for digital communication networks*", New York: McGraw Hill, 1969.

[M81] B.C. Moore, "Principal component analysis in linear systems: controllability, observability and model reduction", *IEEE Trans. on Automatic Control*, Vol. 26, 1981, pp. 17-32.

[M85] Y. Meyer, "Principe d'incertitude, bases hilbertiennes et algebras d'operateurs", *Seminaire Bourbaki.*, No. 662, 1985.

[M88] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *Communications of Pure and Applied Mathematics*, Vol. 41, 1988, pp. 674-693.

[M89] S. G. Mallat, "Multiresolution Approximation and Wavelets," *Trans. American Mathematical Society*, Vol. 315, 1989, pp. 69-88.

[M93] Y. Meyer, "*Wavelets: Algorithms & Applications*", Philadelphia: SIAM, 1993.

[M98] S. Mallat, "*A wavelet tour of signal processing*", 2nd Edition, London: Academic Press, 1998.

[MAF+82] J. Morlet, G. Arens, I. Fourgeau, and D. Giard, "Wave Propagation and Sampling Theory", *Geophysics*, Vol. 47, 1982, pp. 203-236.

[ML03] J.R.L. de Mattos and E.P. Lopes, "A wavelet Galerkin method applied to partial differential equations with variable coefficients", *Fifth Mississippi State Conference on Differential Equations and Computational Simulations, Electronic Journal of Differential Equations*, Conference 10, 2003, pp. 211–225.

[ML91] J.M. Maron and R.J. Lopez, "*Numerical Analysis: A Practical Approach*", 3rd Edition, Belmont: Wadsworth Publishing Company, 1991.

[MM01] G. Miano and A. Maffucci, "Transmission lines and lumped circuits", San Diego: Academic Press, 2001.

[MNZ92] S. Manney, M.S. Nakhla and Q.J. Zhang, "Time domain analysis of nonuniform frequency dependent high-speed interconnects", *Digest of IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Santa Clara, CA, Nov. 1992, pp. 449–453.

[MV03] C. Moler and C. Van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later", *SIAM Review*, Vol. 45, March 2003, pp. 3-49.

[N04] L.W. Nagel, "Is it time for SPICE4?", [Online], June 2004, [Accessed 26 October 2004]. Available: http://www.cs.sandia.gov/nacdm/talks/Nagel_Larry_NACDM2004.ppt

[N69] S. P. Nørsett, "An A-stable modification of the Adams-Bashforth methods", *Lecture Notes in Mathematics,* Vol. 109, 1969, pp. 214-219.

[N75] L.W. Nagel, "SPICE2, A Computer Program to Simulate Semiconductor Circuits", *Tech. Rep. ERL-M520*, Univ. California, Berkeley, May 1975.

[N95] A. R. Newton, "Presentation of the 1995 Phil Kaufman Award to Professor Donald O. Pederson" on 16th November 1995, [Online], [Accessed 10 October 2004]. Available: http://www-cad.eecs.berkeley.edu/~newton/Presentations/DOPPresent.html

[NA02] M. Nakhla and R. Achar, "*The Complete Multimedia Book Series on Signal Integrity (High-Speed Circuit and Interconnect Analysis)*", CD ROM, Nepean, ON, Canada: Omniz Global Knowledge Corporation, 2002.

[NA04] M. Nakhla and R. Achar, "Simulation of distributed interconnects", *Notes on TFB tutorial, International Microwave Symposium (IMS)*, Fort Worth, TX, USA, June 2004.

[NAW04] M. Nakhla, R. Achar and A. Weisshaar, "Fundamentals and trends in modelling and simulation of high-frequency microwave circuits, interconnects and systems", *Notes on TFB tutorial, International Microwave Symposium (IMS)*, Fort Worth, TX, USA, June 2004.

[NL96] E. Ngoya and R. Larcheveque, "Envelope transient analysis: A new method for the transient and steady-state analysis of microwave communications circuits and systems", in *Digest IEEE MTT-S International Microwave Symposium*, San Francisco, CA, USA, Vol. 3, June 1996, pp. 1365–1368.

[NNA96] T. Noda, N. Nagaoka and A. Ametani, "Phase Domain Modelling of Frequency-Dependent Transmission Lines by Means of an ARMA Model", *IEEE Trans. on Power Delivery*, Vol. 11, January 1996, pp. 401–411.

[NNA97] T. Noda, N. Nagaoka and A. Ametani, "Further Improvements to a Phase-Domain ARMA Line Model in Terms of Convolution, Steady-State Initialization, and Stability", *IEEE Trans. on Power Delivery*, Vol. 12, July 1997, pp. 1327–1334.

[NR71] L. Nagel and Rohrer, "Computer analysis of nonlinear circuits, excluding radiation (CANCER)", *IEEE Journal of Solid-State Circuits*, Vol. 6, Aug 1971, pp. 166 - 182.

[NSH+02] R. Neumayer, A. Stelzer, F. Haslinger and R. Weigel, "On the synthesis of equivalent-circuit models for multiports characterized by frequency-dependent parameters", *IEEE Trans. on Microwave Theory and Techniques,* Vol. 50, Dec. 2002, pp. 2789–2796.

[NV76] M. Nakhla and J. Vlach, "A piecewise harmonic balance technique for determination of periodic response of nonlinear systems", *IEEE Trans. on Circuits and Systems*, Vol. 23, Feb. 1976, pp. 85-91.

[O00] K.S. Oh, "Accurate transient simulation of transmission lines with the skin effect", *IEEE Trans. on Computer-aided Design of Integrated Circuits and Systems*, Vol. 19, Mar. 2000, pp. 389-396.

[OCP98] A. Odabasioglu, M. Celik and L.T. Pileggi, "PRIMA: Passive reduced-order interconnect macromodeling algorithm", *IEEE Trans. on Computer-Aided Design*, Vol. 17, Aug. 1998, pp. 645-654.

[OS95] K.S. Oh and J. Schutt-Aine, "Efficient modelling of interconnects and capacitive discontinuities in high-speed digital circuits", *Scientific Report*, University of Illinois, Urbana, USA, June 1995.

[OSB99] A.V. Oppenheim, R.W. Schafer and R. Buck, "*Discrete-Time Signal Processing*", 2nd Edition, New York: Prentice Hall, 1999.

[OWH+96] A.V. Oppenheim, A.S. Willsky and S. H. Nawab, "Signals and Systems", 2nd Edition, New York: Prentice Hall, 1996.

[P00] J.R. Phillips, "Projection frameworks for model reduction of weakly nonlinear systems", in *Proc. Conference on Design Automation (DAC)*, Los Angeles, CA, USA, June 2000, pp. 184-189.

[P03] G.W. Pan. "*Wavelets in electromagnetics and device modelling*", New Jersey: Wiley-Interscience, 2003.

[P04] R. Polikar, "*The wavelet tutorial*", [Online], January 2001, [Accessed on 16 May 2004]. Available: http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html

[P63] D. A. Pope, "An exponential method of numerical integration of ordinary differential equations", *Communications of the ACM*, Vol. 6, 1963, pp. 491-493.

[P94] C.R Paul, "*Analysis of multiconductor transmission lines*", New York: John Willey & Sons, 1994.

[P96] G. Pan, "Orthogonal wavelets with applications in electromagnetics", *IEEE Trans. on Magnetics*, Vol. 32, May 1996, pp. 975 – 983.

[P98] D.M. Pozar, "*Microwave Engineering*", 2nd edition, New York: John Willey & Sons, 1998.

[PB97] P. Perry and T.J. Brazil, "Hilbert Transform Derived Relative Group Delay", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 45, 1997, pp. 1214-1225.

[PB98] P. Perry and T.J. Brazil, "Forcing causality on S-parameter data using the Hilbert Transform", *IEEE Microwave and Guided Letters*, Vol. 8, 1998, pp. 378-380.

[PC02] J.C. Pedro and N.B. Carvalho, "Simulation of RF circuits driven by modulated signals without bandwidth constraints", in *IEEE MTT-S International Microwave Symposium Digest*, Seattle, USA, Vol. 3, June 2002, pp. 2173-2176.

[PDE88] *The Penguin Dictionary of Electronics*, 2nd Edition, The Penguin Group: London, 1988.

[PL89] O. Palusinski and A. Lee, "Analysis of transients in nonuniform and uniform multiconductor transmission lines", *IEEE Trans. on Microwave Theory and Techniques*, Jan. 1989, pp.127–138.

[PR90] L.T. Pillage and R.A Rohrer, "Asymptotic waveform evaluation for timing analysis", *IEEE Trans. on Computer-aided Design*, Vol. 9, April 1990, pp. 352-366.

[R01] J.S. Roychowdhury, "Analysing circuits with widely separated timescales using numerical PDE methods", *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications* [see also IEEE Trans. on Circuits and Systems I: Regular Papers], Vol. 48, May 2001, pp. 578–594.

[R74] A.E. Ruehli, "Equivalent circuit models for three-dimensional multi-conductor systems", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 22, 1974, pp. 216-221.

[RBR92] V. Raghavan, J.E. Bracken and R.A. Rohrer, "AWESpice: a general tool for the accurate and efficient simulation of interconnect problems", in *Proc. 29th ACM/IEEE Design Automation Conference*, Anaheim, CA, USA, June 1992, pp. 87-92.

[RC01] A.E. Ruehli and A.C. Cangellaris, "Progress in the methodologies for the electrical modelling of interconnects and electronic packages", in *Proc. IEEE*, Vol. 89, May 2001, pp. 740–771.

[RN81] R. A. Rohrer and H. Nosrati, "Passivity considerations in stability studies of numerical integration algorithms," *IEEE Trans. on Circuits and Systems*, Vol. 28, Sep. 1981, pp. 857–866.

[RNP94] J.S. Roychowdhury, A.R. Newton, and D.O. Pederson, "Algorithms for the transient simulation of lossy interconnect", *IEEE Trans. on Computer-aided Design of Integrated Circuits and Systems*, Vol. 13, Jan. 1994, pp. 96–104.

[RP91] J.S. Roychowdhury and D.O. Pederson, "Efficient transient simulation of lossy interconnect", in *Proc. 29th ACM/IEEE Design Automation Conference*, San Francisco, CA USA, June 1991, pp. 740-745.

[RP99] P. Rabei and M. Pedram, "Model order reduction of large circuits using balanced truncation" in *Proc. Asia and South Pacific Design Automation Conference* (ASP-DAC), Wanchai, Hong Kong, Jan. 1999, pp.237-240.

[RPH83] J. Rubinstein, P. Penfield, and M. Horowitz, "Signal delay in RC trees", *IEEE Trans. on Computer-Aided Design*, Vol. 2, July 1983, pp. 202–211.

[RW03] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micromachined devices", *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 22, Feb. 2003, pp. 155–170.

[RWG82] S. Rao, D. Wilton and A. Glisson, "Electromagnetic scattering by surfaces of arbitrary shape", *IEEE Trans. on Antennas and Propagation*, Vol. 30, May 1982, pp. 409-418.

[S82] J.O. Stromberg, "A modifed Franklin system and higher order spline systems on R^n as unconditional bases for hardy spaces", in *Conference in honour of A. Zygmund*, Wadsworth Mathematical series, Vol. II, 1982, pp. 475-493.

[S92] M. Shoji, "*Theory of CMOS digital circuits and circuit failures*", Princeton: Princeton University Press, 1992.

[S96] Y. Sheng, "Wavelet Transform", in *The transforms and applications handbook,* A.D. Poularikas, Ed. Boca Raton, USA: CRC Press, 1996, pp. 747-827.

[S96] D. Sharrit, "New method of analysis of communication systems", in Proc. *IEEE MTT-S* WMFA: *Nonlinear CAD Workshop*, San Francisco, USA, June 1996.

[S97] M. Stabrowski, "An efficient algorithm for solving stiff ordinary differential equations', *Simulation Practice and Theory*, Vol. 5, May 1997, pp. 333-344.

[SAN04a] D. Saraswat, R. Achar and M.S. Nakhla, "A fast algorithm and practical considerations for passive macromodeling of measured/simulated data", *IEEE Trans. on Advanced Packaging* [see also IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B: Advanced Packaging],Vol. 27, Feb. 2004, pp. 57-70.

[SAN04b] D. Saraswat, R. Achar and M.S. Nakhla, "Circuit simulation of S-parameter based interconnects via passive macromodels", in *2nd Annual IEEE Northeast Workshop on Circuits and Systems (NEWCAS)*, June 2004, pp. 309-312.

[SCN+94] R. Sanaie, E. Chiprout, M. Nakhla, and Q. J. Zhang, "A fast method for frequency and time domain simulation of high-speed VLSI interconnects," IEEE Trans. on Microwave Theory and Techniques, vol. 42, Dec. 1994, pp. 2562–2571.

[SEW+94] L.M. Silveira, I.M. Elfadel and J.K. White, M. Chilukuri and K.S. Kundert, "An efficient approach to transmission line simulation using measured or tabulated S-parameter data", in *Proc. 31st IEEE/ACM Conference on Design Automation (DAC)*, San Diego, USA, June 1994, pp. 634-639.

[SHP+01] G. Servel, F. Huret, E. Paleczny, J.F. Legier and D. Deschacht, "Inductive effects on crosstalk evaluation", in *Proc. IEEE 2001 International Interconnect Technology Conference*, Burlingame, USA, June 2001, pp. 63-65.

[SKE96] L.M. Silveira, M. Kamen, I. Elfadel, and J. White, "A coordinate-transformed Arnoldi algorithm for generating guaranteed stable reduced-order models for RLC circuits" *Digest IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, San Jose, CA, USA, Nov. 1996, pp. 288-294.

[SM03] W. Schilders and P. Meuris, "Coupled device, circuit and interconnect simulation", in *Proc. 33rd Conference on European Solid-State Device Research (ESSDERC)*, Estoril, Portugal, Sep. 2003, pp. 17–20.

[SM88] J.E. Schutt-Aine and R. Mittra, "Scattering parameter transient analysis of transmission lines loaded with non-linear terminations", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 36, Mar. 1988, pp. 529-536.

[SSW02] T.K. Sarkar, M. Salazar-Palma and M.C. Wicks, "*Wavelet applications in engineering electromagnetics*", Boston: Artech House, 2002.

[SV72] B.K. Swartz and R.S. Varga, "Error bounds for spline and L-spline interpolation", Journal of Approximation Theory, Vol. 6, 1972, pp. 6-49.

[TN92] T. Tang and M. Nakhla, "Analysis of high-speed VLSI interconnect using asymptotic waveform evaluation technique," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 11, Mar. 1992, pp. 341–352.

[V99] C. Valens, "A really friendly guide to wavelets", [Online], 1999, [Accessed 16 May 2004]. Available: http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html

[W04] A. Weisshaar, "High frequency interconnect models", *Notes on TFB tutorial, International Microwave Symposium (IMS)*, Fort Worth, TX, USA, June 2004.

[W63] L.M. Wedepohl, "Application of matrix methods to the solution of travelling wave phenomena in polyphase systems", *Proc. IEE,* Vol. 110, 1963, pp. 2200-2212.

[W98] X-Y. Wu, "A sixth-order A-stable explicit one-step method for stiff systems", *Computers and Mathematics with Applications*, Vol. 35, 1998, pp. 59-64.

[WA94] J.R. Williams and K. Amaratunga, "Simulation Based Design Using Wavelets", Proc. *International Society for Optical Engineering (SPIE) - Wavelet Applications*, Orlando, FA, April 1994, pp. 499-510.

[WC97] D.J. Wilcox and M. Condon, "A new transmission-line model for time-domain implementation", *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering (COMPEL),* Vol. 16, 1997, pp.261-274.

[WCQ+02] J.M. Wang, C-C. Chu; Q. Yu and E.S. Kuh, "On projection-based algorithms for model-order reduction of interconnects", *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications* [see also IEEE Trans. on Circuits and Systems I: Regular Papers], Vol. 49, Nov. 2002, pp. 1563 – 1585.

[WK00] J.M. Wang and E.S. Kuh, "Recent development in interconnect modelling", *in Interconnects in VLSI design,* H. Grabinski, Ed. Boston: Kluwer Academic Publishers, 2000.

[WMS+01] T. Wittig, I. Munteanu, R. Schuhmann and T. Weiland, "Two-step Lanczos algorithm for model order reduction", *IEEE Trans. on Magnetics*, Vol. 38,March 2001, pp. 673-676.

[WT04] *Wavelet Toolbox User's Guide ver. 3.0*, [Online], 2004, [Accessed 24 June 2004]. Available: http://www.mathworks.com/access/helpdesk/help/pdf_doc/wavelet/wavelet_ug.pdf

[WW92] R. Wang and O. Wing, "Transient analysis of dispersive VLSI interconnects terminated in nonlinear loads", *IEEE Trans. on Computer – Aided Design*, vol. 11, Oct. 1992, pp. 1258-1277.

[WYK99] J.M.L. Wang, Q. Yu, and E.S. Kuh, "Coupled noise estimation for distributed RC interconnect model", in *Proc. Design Automation and Test in Europe (DATE)*, Munich, Germany, Mar. 1999, pp. 164-168.

[XLW+00] Q.-W Xu, Z.-F Li, J. Wang and J.-F Mao, "Transient analysis of lossy interconnects by modified method of characteristics", *IEEE Trans. on Circuits and Systems I*, Vol. 47, 2000, pp. 363-375.

[XS92] J. Xu, and W.-C. Shann, "Galerkin-wavelet methods for two-point boundary value problems", *Journal of Numerical Mathematics*, Vol. 63, 1992, pp. 123-144.

[Y90] P.C.L. Yip, "*High frequency circuit design and measurements*", London: Chapman and Hill, 1990.

[Y93] R.K. Young, "*Wavelet theory and its applications*" Boston: Kluwer Academic Publishers, 1993.

[YFW82] C. Yen, Z. Fazarinc and R. L. Wheeler, "Time-domain skin-effect model for transient analysis of lossy transmission lines," *Proc. IEEE*, 1982, pp. 750–757.

[YK95] Q. Yu and E. S. Kuh, "Exact moment-matching model of transmission lines and application to interconnect delay estimation", *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, Vol. 3, Jun. 1995, pp. 311–322.

[YWK99] Q. Yu, J.M. L. Wang, and E.S. Kuh, "Passive multipoint moment matching model order reduction algorithm on multi port distribution interconnect networks", *IEEE Trans. on Circuits and Systems I*, Vol. 46, Jan. 1999, pp. 140-160.

APPENDIX A

Linear algebra

Some linear algebra techniques and terminology employed in this dissertation are summarised in this Appendix. They are taken from [G79] and [D97].

Characteristic polynomial, eigenvalues and eigenvectors

- The polynomial $p(\lambda) = det(\boldsymbol{A} - \lambda \boldsymbol{I})$ is called the *characteristic polynomial* of $\boldsymbol{A}$.
- The roots of $p(\lambda) = 0$ are the *eigenvalues* of $\boldsymbol{A}$.
- A nonzero vector $\boldsymbol{x}$ satisfying

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

is called a (right) *eigenvector* for the eigenvalue λ.

Similarity transform:

Let $\boldsymbol{S}$ be any non-singular matrix. The matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are called *similar matrices* if:

$$\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S},$$

S is a *similarity transformation*. If matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar, they have the same eigenvalues.

Some special matrices

- A square matrix $\boldsymbol{A}$ such that its transpose $\boldsymbol{A}^T = \boldsymbol{A}$ is called *symmetric*.
- A square matrix $\boldsymbol{A}$ such that its transpose conjugate $\boldsymbol{A}^* = \boldsymbol{A}$ is called *Hermitian*.
- A real symmetric (complex Hermitian) matrix $\boldsymbol{A}$ is *positive definite* if

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0 \ (\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x} > 0), \quad \forall \boldsymbol{x} \neq 0.$$

Orthogonal matrices

A real matrix $\boldsymbol{Q}$ is *orthogonal* if

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}.$$

An orthogonal matrix has the following properties:

1. All columns, $\boldsymbol{q}_i$ of orthogonal matrices have unit two norms:

   $$\|\boldsymbol{q}_i\|_2 = 1,$$

   which implies that

   $$\boldsymbol{q}_i^T\boldsymbol{q}_i = 1.$$

2. All columns, $\boldsymbol{q}_i$ of orthogonal matrices are orthogonal to each other

   $$\boldsymbol{q}_i^T\boldsymbol{q}_j = 0 \quad (i \neq j).$$

3. If $\boldsymbol{Q}$ is square matrix, then

   $$\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T.$$

Orthonormal matrices

A complex matrix $\boldsymbol{Q}$ is *orthonormal* if

$$\boldsymbol{q}_i^*\boldsymbol{q}_j = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

If $\boldsymbol{Q}$ is a real matrix then the orthonormality condition reduces to

$$\boldsymbol{q}_i^T\boldsymbol{q}_j = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

QR decomposition

Let $\boldsymbol{K}$ be an *mxn* matrix with $m>n$ and with full column rank. Then there exists a unique *mxn* orthogonal matrix $\boldsymbol{Q}$ and a unique upper-triangular matrix $\boldsymbol{R}_u$ with positive diagonals $(r_{ii}>0)$ such that

$$\boldsymbol{K} = \boldsymbol{Q}\boldsymbol{R}_u$$

There are several techniques available to perform orthogonalization. The most widely used is the modified Gram-Schmidt orthogonalization process.

LU factorisation

LU factorisation is the procedure for decomposing a square matrix $\boldsymbol{A}$ of order n into a product of a *lower triangular matrix* $\boldsymbol{L}$ and an *upper triangular matrix* $\boldsymbol{U}$, i.e.

$$\boldsymbol{A} = \boldsymbol{LU}$$

It is used to solve the matrix equation:

$$\boldsymbol{Ax} = \boldsymbol{b},$$

since

$$\boldsymbol{Ax} = (\boldsymbol{LU})\boldsymbol{x} = \boldsymbol{L}(\boldsymbol{Ux}) = \boldsymbol{b}$$

Using forward substitution, the intermediate vector $\boldsymbol{y}$ is found from:

$$\boldsymbol{Ly} = \boldsymbol{b}$$

and then, using backward substitution for the required solution $\boldsymbol{x}$ is found as

$$\boldsymbol{Ux} = \boldsymbol{y}.$$

Upper-Hessenberg matrix

A matrix $\boldsymbol{H}$ is called *upper-Hessenberg* if $H_{ij}=0$ for $(i>j+1)$. For example, consider an upper Hessenberg matrix of order n, having the following, so called, *companion form*

$$\boldsymbol{H} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & -c_1 \\ 1 & 0 & 0 & \cdots & 0 & -c_2 \\ 0 & 1 & 0 & \cdots & 0 & -c_3 \\ 0 & 0 & 1 & \cdots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & 0 & -c_{n-1} \\ 0 & 0 & 0 & \cdots & 1 & -c_n \end{bmatrix}$$

The characteristic polynomial $\rho(x)$ for the Hessenberg matrix in companion form may be analytically computed as:

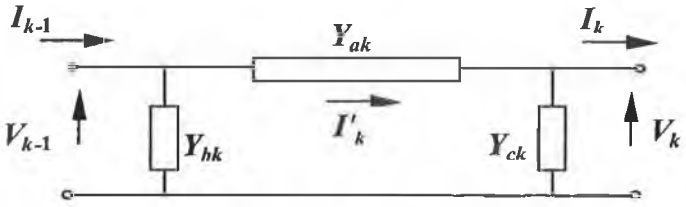
$$\rho(x) = x^n + \sum_{i=1}^{n} c_i x^{i-1}.$$

The roots of $\rho(x)$ give the eigenvalues of $\boldsymbol{H}$.

APPENDIX B

The ABCD matrices for the resonant model

The ***ABCD*** matrices for the resonant model can be expressed directly in terms of Z_{ak}, Z_{bk} and Z_{ck} defined in the equivalent-π representation of the k^{th} section



Equivalent-π representation of k^{th} section

Matrix *A*

The matrix $\boldsymbol{A}$ is defined as

$$\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2 \tag{B.1}$$

where

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{Y}_A & -\boldsymbol{Y}_A \\ -\boldsymbol{Y}_A & \boldsymbol{Y}_A \end{bmatrix}$$
$$\boldsymbol{A}_2 = \begin{bmatrix} \boldsymbol{Y}_{b1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_{cK} \end{bmatrix} \tag{B.2}$$

In (B.2), $\boldsymbol{Y}_A$ corresponds to the total series impedance:

$$\boldsymbol{Y}_A = \left[\sum_{k=1}^{K} \boldsymbol{Z}_{ak} \right]^{-1}. \tag{B.3}$$

Matrix *C*

The matrix $\boldsymbol{C}$ is given by

$$\boldsymbol{C} = [\boldsymbol{c}_1 \;\; \boldsymbol{c}_2] = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ . & . \\ . & . \\ c_{K-1,1} & c_{K-1,2} \end{bmatrix}, \tag{B.4}$$

where

$$c_{i1} = \sum_{k=i+1}^{K} \boldsymbol{Z}_{ak} \boldsymbol{Y}_A \qquad \text{and} \qquad c_{i2} = \sum_{k=1}^{i} \boldsymbol{Z}_{ak} \boldsymbol{Y}_A \,. \tag{B.5}$$

Matrix *B*

B is then given by :

$$\boldsymbol{B} = \boldsymbol{C}^T \boldsymbol{Y} \,, \tag{B.6}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{Y}_{K-1} \end{bmatrix} \tag{B.7}$$

in which

$$\boldsymbol{Y}_k = \boldsymbol{Y}_{ck} + \boldsymbol{Y}_{b,k+1} \,. \tag{B.8}$$

Matrix *D*

Finally, the square matrix ***D*** is specified by

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{d}_{11} & \boldsymbol{d}_{12} & \cdots & \boldsymbol{d}_{1,K-1} \\ \boldsymbol{d}_{21} & \boldsymbol{d}_{22} & \cdots & \boldsymbol{d}_{2,K-1} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{d}_{K-1,1} & \boldsymbol{d}_{K-1,2} & \cdots & \boldsymbol{d}_{K-1,K-1} \end{bmatrix}, \tag{B.9}$$

where

$$\begin{aligned} \boldsymbol{d}_{ij} &= -\left[\sum_{p=i+1}^{K} \boldsymbol{Z}_{ap}\right]\left[\boldsymbol{Y}_A\right]\left[\sum_{q=1}^{j} \boldsymbol{Z}_{aq}\right]\left[\boldsymbol{Y}_j\right], && \text{for } i \geq j \\ \boldsymbol{d}_{ij} &= -\left\{\left[\sum_{p=i+1}^{K} \boldsymbol{Z}_{ap}\right]\left[\boldsymbol{Y}_A\right]\left[\sum_{q=1}^{j} \boldsymbol{Z}_{aq}\right] - \left[\sum_{p=i+1}^{j} \boldsymbol{Z}_{ap}\right]\right\}\left[\boldsymbol{Y}_j\right], && \text{for } j > i \end{aligned} \tag{B.10}$$

In the important particular case of a uniform transmission line divided into K sections of equal length l, the formulae for the submatrices of ***D*** simplify to

$$d_{ij} = -Z_a Y_{bc}\left(\frac{K-i}{K}\right) j \qquad \text{for } i \geq j$$
$$d_{ij} = -Z_a Y_{bc}\left(\frac{K-j}{K}\right) i \qquad \text{for } j > i \tag{B.11}$$

where

$$Z_a = \sinh(\Gamma l) Z_0 \tag{B.12}$$

$$Y_{bc} = 2Y_0 \tanh\left(\frac{\Gamma l}{2}\right) . \tag{B.13}$$

For a lossless line, the $\boldsymbol{D}$ matrix can further be simplified to the following form

$$\boldsymbol{D} = 4\sin^2\left(\frac{\omega l\sqrt{LC}}{2}\right)\boldsymbol{R} . \tag{B.14}$$

APPENDIX C

The history currents $\boldsymbol{i}_{his1}$ and $\boldsymbol{i}_{his2}$

In this Appendix, the complete procedure for translating the $\mathcal{Z}$-domain line model written as in (C.1) into the time domain form (C.2) is presented.

$$\begin{bmatrix} \boldsymbol{I}_S(z) \\ -\boldsymbol{I}_R(z) \end{bmatrix} = [\boldsymbol{Y}_B(z)] \begin{bmatrix} \boldsymbol{V}_S(z) \\ \boldsymbol{V}_R(z) \end{bmatrix} \tag{C.1}$$

$$\begin{bmatrix} \boldsymbol{i}_S \\ -\boldsymbol{i}_R \end{bmatrix}^{(r)} = [\boldsymbol{y}_B]^{(r)} \begin{bmatrix} \boldsymbol{v}_S \\ \boldsymbol{v}_R \end{bmatrix}^{(r)} + \begin{bmatrix} \boldsymbol{i}_{his1} \\ \boldsymbol{i}_{his2} \end{bmatrix}^{(r-1)} \tag{C.2}$$

The superscript 'r' denotes values at the time t_r.

From (4.55), the expression for matrix $\boldsymbol{Y}_B(z)$ is given as:

$$\boldsymbol{Y}_B(z) = \boldsymbol{Y}_b(z) + \boldsymbol{Y}_{BB}^{*}(z) + \boldsymbol{P}\boldsymbol{\zeta}\boldsymbol{g}\boldsymbol{P}^T(z). \tag{C.3}$$

From (C.1) and (C.3) it follows:

$$\begin{bmatrix} \boldsymbol{I}_S(z) \\ -\boldsymbol{I}_R(z) \end{bmatrix} = \boldsymbol{Y}_b(z) \begin{bmatrix} \boldsymbol{V}_S(z) \\ \boldsymbol{V}_R(z) \end{bmatrix} + \boldsymbol{Y}_{BB}^{*}(z) \begin{bmatrix} \boldsymbol{V}_S(z) \\ \boldsymbol{V}_R(z) \end{bmatrix} + \boldsymbol{P}\boldsymbol{\zeta}\boldsymbol{g}\boldsymbol{P}^T(z) \begin{bmatrix} \boldsymbol{V}_S(z) \\ \boldsymbol{V}_R(z) \end{bmatrix} \tag{C.4}$$

Equation (C.4) can be more compactly written as

$$\boldsymbol{I}_B(z) = \left[\boldsymbol{Y}_b(z) + \boldsymbol{Y}_{BB}^{*}(z) + \boldsymbol{P}\boldsymbol{\zeta}\boldsymbol{g}\boldsymbol{P}^T(z)\right]\boldsymbol{V}_B(z) = \boldsymbol{I}_b(z) + \boldsymbol{I}_{BB}^{*}(z) + \boldsymbol{I}_{PZG}(z), \tag{C.5}$$

where

$$\boldsymbol{I}_B = \begin{bmatrix} \boldsymbol{I}_S \\ -\boldsymbol{I}_R \end{bmatrix} \quad \text{and} \quad \boldsymbol{V}_B = \begin{bmatrix} \boldsymbol{V}_S \\ \boldsymbol{V}_R \end{bmatrix}. \tag{C.6}$$

The derivation of the time-domain representation will now proceed separately for the three terms in (C.5) given as:

$$\begin{aligned} \boldsymbol{I}_b(z) &= \boldsymbol{Y}_b(z)\boldsymbol{V}_B(z) \\ \boldsymbol{I}_{BB}^{''}(z) &= \boldsymbol{Y}_{BB}^{*}(z)\boldsymbol{V}_B(z) \\ \boldsymbol{I}_{PZG}(z) &= \boldsymbol{P}\boldsymbol{\zeta}\boldsymbol{g}\boldsymbol{P}^T(z)\boldsymbol{V}_B(z) \end{aligned} \tag{C.7}$$

Term $I_b(z) = Y_b(z)V_B(z)$

From (4.40), the matrix $\boldsymbol{Y}_b$ is given as:

$$\boldsymbol{Y}_b = \begin{bmatrix} \frac{1}{KZ_a} & -\frac{1}{KZ_a} \\ -\frac{1}{KZ_a} & \frac{1}{KZ_a} \end{bmatrix}. \tag{C.8}$$

Since the term $\frac{1}{KZ_a}$ is approximated with $\frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}}$ (Section 4.2.2), equation (C.8) becomes:

$$\boldsymbol{Y}_b(z) = \begin{bmatrix} \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} & -\frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \\ -\frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} & \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \end{bmatrix}. \tag{C.9}$$

Therefore, the following matrix equation may be written:

$$\begin{bmatrix} \boldsymbol{I}_S^b(z) \\ -\boldsymbol{I}_R^b(z) \end{bmatrix} = \begin{bmatrix} \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} & -\frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \\ -\frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} & \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_S(z) \\ \boldsymbol{V}_R(z) \end{bmatrix}, \tag{C.10}$$

i.e.

$$\boldsymbol{I}_S^b(z) = \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \boldsymbol{V}_S(z) - \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \boldsymbol{V}_R(z) \tag{C.11}$$

$$-\boldsymbol{I}_R^b(z) = -\frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \boldsymbol{V}_S(z) + \frac{a_1^b z^{-1} + a_2^b z^{-2}}{1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}} \boldsymbol{V}_R(z) \tag{C.12}$$

Cross multiplying, the equation (C.11) becomes

$$\left(1 + b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}\right) \boldsymbol{I}_S^b(z) = \left(a_1^b z^{-1} + a_2^b z^{-2}\right)\left(\boldsymbol{V}_S(z) - \boldsymbol{V}_R(z)\right) \tag{C.13}$$

or,

$$\boldsymbol{I}_S^b(z) = \left(a_1^b z^{-1} + a_2^b z^{-2}\right)\left(\boldsymbol{V}_S(z) - \boldsymbol{V}_R(z)\right) - \left(b_1^b z^{-1} + b_2^b z^{-2} + b_3^b z^{-3}\right) \boldsymbol{I}_S^b(z). \tag{C.14}$$

Consider the following property:

$$\alpha z^{-k} X(z) \leftrightarrow \alpha x(r-k). \tag{C.15}$$

Thus, equation (C.14) translates to:

$$i_s^b(r) = a_1^b(v_S - v_R)(r-1) + a_2^b(v_S - v_R)(r-2) - b_1^b i_s^b(r-1) - b_2^b i_s^b(r-2) - b_3^b i_s^b(r-3) \tag{C.16}$$

Similarly, equation (C.14) translates to:

$$-i_R^b(r) = -a_1^b(v_S - v_R)(r-1) - a_2^b(v_S - v_R)(r-2) + b_1^b i_R^b(r-1) + b_2^b i_R^b(r-2) + b_3^b i_R^b(r-3). \tag{C.17}$$

In matrix form, equations (C.16) and (C.17) may be written as:

$$\begin{bmatrix} i_S^b \\ -i_R^b \end{bmatrix}^{(r)} = \begin{bmatrix} a_1^b & -a_1^b \\ -a_1^b & a_1^b \end{bmatrix} \begin{bmatrix} v_S \\ v_R \end{bmatrix}^{(r-1)} + \begin{bmatrix} a_2^b & -a_2^b \\ -a_2^b & a_2^b \end{bmatrix} \begin{bmatrix} v_S \\ v_R \end{bmatrix}^{(r-2)} + \begin{bmatrix} -b_1^b & 0 \\ 0 & -b_1^b \end{bmatrix} \begin{bmatrix} i_S^b \\ -i_R^b \end{bmatrix}^{(r-1)} + \begin{bmatrix} -b_2^b & 0 \\ 0 & -b_2^b \end{bmatrix} \begin{bmatrix} i_S^b \\ -i_R^b \end{bmatrix}^{(r-2)} + \begin{bmatrix} -b_3^b & 0 \\ 0 & -b_3^b \end{bmatrix} \begin{bmatrix} i_S^b \\ -i_R^b \end{bmatrix}^{(r-3)} \tag{C.18}$$

Since all the elements on the right hand side depend only on past values, it is possible to write:

$$\begin{bmatrix} i_S^b \\ -i_R^b \end{bmatrix}^{(r)} = \begin{bmatrix} i_{his1}^b \\ i_{his2}^b \end{bmatrix}^{(r-1)}, \tag{C.19}$$

where:

$$\begin{bmatrix} i_{his1}^b \\ i_{his2}^b \end{bmatrix}^{(r-1)} = \begin{bmatrix} a_1^b v_S^b - a_1^b v_R^b - b_1^b i_S^b \\ -a_1^b v_S^b + a_1^b v_R^b - b_1^b \left(-i_R^b\right) \end{bmatrix}^{(r-1)} + \begin{bmatrix} a_2^b v_S^b - a_2^b v_R^b - b_2^b i_S^b \\ -a_2^b v_S^b + a_2^b v_R^b - b_2^b \left(-i_S^b\right) \end{bmatrix}^{(r-2)} + \begin{bmatrix} -b_3^b i_S^b \\ -b_3^b \left(-i_R^b\right) \end{bmatrix}^{(r-3)} \tag{C.20}$$

Term $I_{BB}''(z) = Y_{BB}''(z) V_B(z)$

From (4.40) the matrix Y_{BB}'' is given as:

$$Y_{BB}'' = \begin{bmatrix} \frac{Y_{bb}}{2} & 0 \\ 0 & \frac{Y_{bb}}{2} \end{bmatrix}. \tag{C.21}$$

Since the term $Y_{bb}/2$ is approximated with $\frac{a_0^{BB} + a_1^{BB} z^{-1}}{1 + b_1^{BB} z^{-1}}$ (Section 4.2.2), equation (C.21) becomes:

$$Y''_{BB}(z)=\begin{bmatrix} \dfrac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}} & 0 \\ 0 & \dfrac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}} \end{bmatrix} \tag{C.22}$$

Therefore, the following matrix equation may be written:

$$\begin{bmatrix} I_S^{BB}(z) \\ -I_R^{BB}(z) \end{bmatrix}=\begin{bmatrix} \dfrac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}} & 0 \\ 0 & \dfrac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}} \end{bmatrix}\begin{bmatrix} V_S^{BB}(z) \\ V_R^{BB}(z) \end{bmatrix} \tag{C.23}$$

that is,

$$I_S^{BB}(z)=\frac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}}V_S^{BB}(z) \tag{C.24}$$

$$-I_R^{BB}(z)=\frac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}}V_R^{BB}(z). \tag{C.25}$$

As before:

$$\begin{bmatrix} i_S^{BB} \\ -i_R^{BB} \end{bmatrix}^{(r)}=\begin{bmatrix} a_0^{BB} & 0 \\ 0 & a_0^{BB} \end{bmatrix}\begin{bmatrix} v_S^{BB} \\ v_R^{BB} \end{bmatrix}^{(r)}+\begin{bmatrix} a_1^{BB} & 0 \\ 0 & a_1^{BB} \end{bmatrix}\begin{bmatrix} v_S^{BB} \\ v_R^{BB} \end{bmatrix}^{(r-1)}+\begin{bmatrix} -b_1^{BB} & 0 \\ 0 & -b_1^{BB} \end{bmatrix}\begin{bmatrix} i_S^{BB} \\ -i_R^{BB} \end{bmatrix}^{(r-1)} \tag{C.26}$$

Equation (C.26) may be compactly noted as:

$$\begin{bmatrix} i_S^{BB} \\ -i_R^{BB} \end{bmatrix}^{(r)}=\begin{bmatrix} a_0^{BB} & 0 \\ 0 & a_0^{BB} \end{bmatrix}\begin{bmatrix} v_S^{BB} \\ v_R^{BB} \end{bmatrix}^{(r)}+\begin{bmatrix} i_{his1}^{BB} \\ i_{his2}^{BB} \end{bmatrix}^{(r-1)}, \tag{C.27}$$

where

$$\begin{bmatrix} i_{his1}^{BB} \\ i_{his2}^{BB} \end{bmatrix}^{(r-1)}=\begin{bmatrix} a_1^{BB}v_S^{BB}-b_1^{BB}i_S^{BB} \\ a_1^{BB}v_R^{BB}-b_1^{BB}\left(-i_R^{BB}\right) \end{bmatrix}^{(r-1)}. \tag{C.28}$$

Term $I_{PZG}(z) = P\zeta gP^T(z)V_B(z)$

Considering that ζ and g are diagonal matrices, the matrix $P\zeta gP^T(z)$ may be written as:

$$P\zeta gP^T(z) = \sum_{i=1}^{K-1} \zeta_i(z) g_i(z) p_i p_i^T , \tag{C.29}$$

where K is the number of modes and p_i is the i^{th} column of P. Elements of matrix P are constant, i.e. independent of frequency and they are not approximated. ζ_i and g_i are the i^{th} diagonal element of matrices ζ and g respectively. These diagonal elements are fitted with:

$$\zeta_i = \frac{a_0^\zeta + a_1^\zeta z^{-1}}{1 + b_1^\zeta z^{-1}} \tag{C.30}$$

$$g_i = \frac{a_1^g z^{-1} + a_2^g z^{-2}}{1 + b_1^g z^{-1} + b_2^g z^{-2} + b_3^g z^{-3}} . \tag{C.31}$$

Since they are multiplied it is possible to write:

$$\zeta_i g_i = \frac{a_0^\zeta + a_1^\zeta z^{-1}}{1 + b_1^\zeta z^{-1}} \times \frac{a_1^g z^{-1} + a_2^g z^{-2}}{1 + b_1^g z^{-1} + b_2^g z^{-2} + b_3^g z^{-3}} = \frac{c_1^i z^{-1} + c_2^i z^{-2} + c_3^i z^{-3}}{1 + d_1^i z^{-1} + d_2^i z^{-2} + d_3^i z^{-3} + d_4^i z^{-3}} \tag{C.32}$$

where :

$$\begin{aligned} c_1^i &= a_0^\zeta a_1^g & d_1^i &= b_1^\zeta + b_1^g \\ c_2^i &= a_0^\zeta a_2^g + a_1^\zeta a_1^g & d_2^i &= b_2^g + b_1^\zeta b_1^g \\ c_3^i &= a_1^\zeta a_2^g & d_3^i &= b_3^g + b_1^\zeta b_2^g \\ & & d_4^i &= b_1^\zeta b_3^g \end{aligned} \tag{C.33}$$

Therefore,

$$I_{PZG}(z) = P\zeta gP^T(z)V_B(z) = \sum_{i=1}^{K-1} \zeta_i(z) g_i(z) p_i p_i^T V_B(z) . \tag{C.34}$$

In case of a single line (2 ports), $p_i = \begin{bmatrix} p_1^i \\ p_2^i \end{bmatrix}$ and hence

$$p_i p_i^T = \begin{bmatrix} p_1^i p_1^i & p_1^i p_2^i \\ p_1^i p_2^i & p_2^i p_2^i \end{bmatrix} . \tag{C.35}$$

Equation (C.7) now becomes

$$\begin{aligned} I_{PZG}(z) &= \sum_{i=1}^{K-1} \frac{c_1^i z^{-1} + c_2^i z^{-2} + c_3^i z^{-3}}{1 + d_1^i z^{-1} + d_2^i z^{-2} + d_3^i z^{-3} + d_4^i z^{-3}} \begin{bmatrix} p_1^i p_1^i & p_1^i p_2^i \\ p_1^i p_2^i & p_2^i p_2^i \end{bmatrix} \begin{bmatrix} V_S(z) \\ V_R(z) \end{bmatrix} = \\ &= \sum_{i=1}^{K-1} \frac{c_1^i z^{-1} + c_2^i z^{-2} + c_3^i z^{-3}}{1 + d_1^i z^{-1} + d_2^i z^{-2} + d_3^i z^{-3} + d_4^i z^{-3}} \begin{bmatrix} p_1^i p_1^i V_S(z) + p_1^i p_2^i V_R(z) \\ p_1^i p_2^i V_S(z) + p_2^i p_2^i V_R(z) \end{bmatrix} = \sum_{i=1}^{K-1} \begin{bmatrix} I_S^i(z) \\ -I_R^i(z) \end{bmatrix} \end{aligned} \tag{C.36}$$

where, for $i = 1,...,7$,

$$\boldsymbol{I}_S^i(z) = \frac{c_1^i z^{-1} + c_2^i z^{-2} + c_3^i z^{-3}}{1 + d_1^i z^{-1} + d_2^i z^{-2} + d_3^i z^{-3} + d_4^i z^{-3}} \times p_1^i \left[p_1^i \boldsymbol{V}_S(z) + p_2^i \boldsymbol{V}_R(z) \right] \tag{C.37}$$

$$-\boldsymbol{I}_R^i(z) = \frac{c_1^i z^{-1} + c_2^i z^{-2} + c_3^i z^{-3}}{1 + d_1^i z^{-1} + d_2^i z^{-2} + d_3^i z^{-3} + d_4^i z^{-3}} \times p_2^i \left[p_1^i \boldsymbol{V}_S(z) + p_2^i \boldsymbol{V}_R(z) \right]. \tag{C.38}$$

In addition, in the $\mathcal{Z}$-domain matrix $\boldsymbol{P\zeta gP}^{\mathrm{T}}(z)$ may be noted as:

$$\boldsymbol{P\zeta gP}^{\mathrm{T}}(z) = \sum_{i=1}^{K-1} \frac{c_1^i z^{-1} + c_2^i z^{-2} + c_3^i z^{-3}}{1 + d_1^i z^{-1} + d_2^i z^{-2} + d_3^i z^{-3} + d_4^i z^{-3}} \begin{bmatrix} p_1^i p_1^i & p_1^i p_2^i \\ p_1^i p_2^i & p_2^i p_2^i \end{bmatrix} \tag{C.39}$$

Now applying property (C.15)

$$\begin{aligned} \boldsymbol{i}_S^i(r) = p_1^i \Big[& c_1^i \left(p_1^i \boldsymbol{v}_S(r-1) + p_2^i \boldsymbol{v}_R(r-1) \right) + c_2^i \left(p_1^i \boldsymbol{v}_S(r-2) + p_2^i \boldsymbol{v}_R(r-2) \right) + c_3^i \left(p_1^i \boldsymbol{v}_S(r-3) + \right. \\ & \left. p_2^i \boldsymbol{v}_R(r-3) \right) \Big] - d_1^i \boldsymbol{i}_S^i(r-1) - d_2^i \boldsymbol{i}_S^i(r-2) - d_3^i \boldsymbol{i}_S^i(r-3) - d_4^i \boldsymbol{i}_S^i(r-4) \\ -\boldsymbol{i}_R^i(r) = p_2^i \Big[& c_1^i \left(p_1^i \boldsymbol{v}_S(r-1) + p_2^i \boldsymbol{v}_R(r-1) \right) + c_2^i \left(p_1^i \boldsymbol{v}_S(r-2) + p_2^i \boldsymbol{v}_R(r-2) \right) + c_3^i \left(p_1^i \boldsymbol{v}_S(r-3) + \right. \\ & \left. p_2^i \boldsymbol{v}_R(r-3) \right) \Big] - d_1^i \boldsymbol{i}_R^i(r-1) - d_2^i \boldsymbol{i}_R^i(r-2) - d_3^i \boldsymbol{i}_R^i(r-3) - d_4^i \boldsymbol{i}_R^i(r-4) \end{aligned} \tag{C.40}$$

From (C.34) and (C.36) follows that:

$$\begin{aligned} \boldsymbol{i}_S^{PZG}(r) &= \sum_{i=1}^{K-1} \boldsymbol{i}_S^i(r) \\ -\boldsymbol{i}_R^{PZG}(r) &= \sum_{i=1}^{K-1} \left(-\boldsymbol{i}_R^i(r) \right) \end{aligned} \tag{C.41}$$

Since all the elements on the right hand side depend only on past values, one finally may write

$$\begin{bmatrix} \boldsymbol{i}_S^{PZG} \\ -\boldsymbol{i}_R^{PZG} \end{bmatrix}^{(r)} = \begin{bmatrix} \boldsymbol{i}_{his1}^{PZG} \\ \boldsymbol{i}_{his2}^{PZG} \end{bmatrix}^{(r-1)}, \tag{C.42}$$

where:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{i}_{his1}^{PZG} \\ \boldsymbol{i}_{his2}^{PZG} \end{bmatrix}^{(r-1)} = \sum_{i=1}^{K-1} \Bigg\{ & \begin{bmatrix} c_1^i p_1^i \left(p_1^i \boldsymbol{v}_S + p_2^i \boldsymbol{v}_R \right) - d_1^i \boldsymbol{i}_S^i \\ c_1^i p_2^i \left(p_1^i \boldsymbol{v}_S + p_2^i \boldsymbol{v}_R \right) - d_1^i \left(-\boldsymbol{i}_R^i \right) \end{bmatrix}^{(r-1)} + \begin{bmatrix} c_2^i p_1^i \left(p_1^i \boldsymbol{v}_S + p_2^i \boldsymbol{v}_R \right) - d_2^i \boldsymbol{i}_S^i \\ c_2^i p_2^i \left(p_1^i \boldsymbol{v}_S + p_2^i \boldsymbol{v}_R \right) - d_2^i \left(-\boldsymbol{i}_R^i \right) \end{bmatrix}^{(r-2)} \\ & + \begin{bmatrix} c_3^i p_1^i \left(p_1^i \boldsymbol{v}_S + p_2^i \boldsymbol{v}_R \right) - d_3^i \boldsymbol{i}_S^i \\ c_3^i p_2^i \left(p_1^i \boldsymbol{v}_S + p_2^i \boldsymbol{v}_R \right) - d_3^i \left(-\boldsymbol{i}_R^i \right) \end{bmatrix}^{(r-3)} + \begin{bmatrix} -d_4^i \boldsymbol{i}_S^i \\ -d_4^i \left(-\boldsymbol{i}_R^i \right) \end{bmatrix}^{(r-4)} \Bigg\} \end{aligned} \tag{C.43}$$

Total history currents

From (C.20), (C.28) and (C.43), applying the superposition principle, the following expression for history currents is obtained.

$$\begin{bmatrix} \boldsymbol{i}_{his1} \\ \boldsymbol{i}_{his2} \end{bmatrix}^{(r-1)} = \begin{bmatrix} \boldsymbol{i}^{b}_{his1} \\ \boldsymbol{i}^{b}_{his2} \end{bmatrix}^{(r-1)} + \begin{bmatrix} \boldsymbol{i}^{BB}_{his1} \\ \boldsymbol{i}^{BB}_{his2} \end{bmatrix}^{(r-1)} + \begin{bmatrix} \boldsymbol{i}^{PZG}_{his1} \\ \boldsymbol{i}^{PZG}_{his2} \end{bmatrix}^{(r-1)} =$$

$$\begin{bmatrix} a^{b}_{1}\boldsymbol{v}^{b}_{S} - a^{b}_{1}\boldsymbol{v}^{b}_{R} - b^{b}_{1}\boldsymbol{i}^{b}_{S} \\ -a^{b}_{1}\boldsymbol{v}^{b}_{S} + a^{b}_{1}\boldsymbol{v}^{b}_{R} - b^{b}_{1}\left(-\boldsymbol{i}^{b}_{R}\right) \end{bmatrix}^{(r-1)} + \begin{bmatrix} a^{BB}_{1}\boldsymbol{v}^{BB}_{S} - b^{BB}_{1}\boldsymbol{i}^{BB}_{S} \\ a^{BB}_{1}\boldsymbol{v}^{BB}_{R} - b^{BB}_{1}\left(-\boldsymbol{i}^{BB}_{R}\right) \end{bmatrix}^{(r-1)} + \begin{bmatrix} \sum_{i=1}^{K-1}\left[c^{i}_{1}p^{i}_{1}\left(p^{i}_{1}\boldsymbol{v}_{S} + p^{i}_{2}\boldsymbol{v}_{R}\right) - d^{i}_{1}\boldsymbol{i}^{i}_{S}\right] \\ \sum_{i=1}^{K-1}\left[c^{i}_{1}p^{i}_{2}\left(p^{i}_{1}\boldsymbol{v}_{S} + p^{i}_{2}\boldsymbol{v}_{R}\right) - d^{i}_{1}\left(-\boldsymbol{i}^{i}_{R}\right)\right] \end{bmatrix}^{(r-1)}$$

$$+ \begin{bmatrix} a^{b}_{2}\boldsymbol{v}^{b}_{S} - a^{b}_{2}\boldsymbol{v}^{b}_{R} - b^{b}_{2}\boldsymbol{i}^{b}_{S} \\ -a^{b}_{2}\boldsymbol{v}^{b}_{S} + a^{b}_{2}\boldsymbol{v}^{b}_{R} - b^{b}_{2}\left(-\boldsymbol{i}^{b}_{S}\right) \end{bmatrix}^{(r-2)} + \begin{bmatrix} \sum_{i=1}^{K-1}\left[c^{i}_{2}p^{i}_{1}\left(p^{i}_{1}\boldsymbol{v}_{S} + p^{i}_{2}\boldsymbol{v}_{R}\right) - d^{i}_{2}\boldsymbol{i}^{i}_{S}\right] \\ \sum_{i=1}^{K-1}\left[c^{i}_{2}p^{i}_{2}\left(p^{i}_{1}\boldsymbol{v}_{S} + p^{i}_{2}\boldsymbol{v}_{R}\right) - d^{i}_{2}\left(-\boldsymbol{i}^{i}_{R}\right)\right] \end{bmatrix}^{(r-2)} +$$

$$+ \begin{bmatrix} -b^{b}_{3}\boldsymbol{i}^{b}_{S} \\ -b^{b}_{3}\left(-\boldsymbol{i}^{b}_{R}\right) \end{bmatrix}^{(r-3)} + \begin{bmatrix} \sum_{i=1}^{K-1}\left[c^{i}_{3}p^{i}_{1}\left(p^{i}_{1}\boldsymbol{v}_{S} + p^{i}_{2}\boldsymbol{v}_{R}\right) - d^{i}_{3}\boldsymbol{i}^{i}_{S}\right] \\ \sum_{i=1}^{K-1}\left[c^{i}_{3}p^{i}_{2}\left(p^{i}_{1}\boldsymbol{v}_{S} + p^{i}_{2}\boldsymbol{v}_{R}\right) - d^{i}_{3}\left(-\boldsymbol{i}^{i}_{R}\right)\right] \end{bmatrix}^{(r-3)} + \begin{bmatrix} -d^{i}_{4}\boldsymbol{i}^{i}_{S} \\ -d^{i}_{4}\left(-\boldsymbol{i}^{i}_{R}\right) \end{bmatrix}^{(r-4)}$$

(C.44)

In addition, considering (C.19), (C.27) and (C.42), equation (C.5) may now be written in the time domain as:

$$\begin{bmatrix} \boldsymbol{i}_{S} \\ -\boldsymbol{i}_{R} \end{bmatrix}^{(r)} = \begin{bmatrix} \boldsymbol{i}^{b}_{S} \\ -\boldsymbol{i}^{b}_{R} \end{bmatrix}^{(r)} + \begin{bmatrix} \boldsymbol{i}^{BB}_{S} \\ -\boldsymbol{i}^{BB}_{R} \end{bmatrix}^{(r)} + \begin{bmatrix} \boldsymbol{i}^{PZG}_{S} \\ -\boldsymbol{i}^{PZG}_{R} \end{bmatrix}^{(r)} =$$

$$= \begin{bmatrix} a^{BB}_{0} & 0 \\ 0 & a^{BB}_{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}^{BB}_{S} \\ \boldsymbol{v}^{BB}_{R} \end{bmatrix}^{(r)} + \begin{bmatrix} \boldsymbol{i}^{b}_{his1} + \boldsymbol{i}^{BB}_{his1} + \boldsymbol{i}^{PZG}_{his1} \\ \boldsymbol{i}^{b}_{his2} + \boldsymbol{i}^{BB}_{his2} + \boldsymbol{i}^{PZG}_{his2} \end{bmatrix}^{(r-1)}$$

(C.45)

Comparing (C.45) and (C.2) it is clear that:

$$\left[\boldsymbol{y}_{B}\right]^{(r)} = \begin{bmatrix} a^{BB}_{0} & 0 \\ 0 & a^{BB}_{0} \end{bmatrix}. \tag{C.46}$$

APPENDIX D

The choice of the approximating order for the ARMA functions

In this Appendix, the reasoning behind the choice of the order of the approximating ARMA functions is given by studying the nature of the frequency-dependant elements of the resonant model for the case of a lossless line as presented in [C98].

For a *lossless* line, the elements of $\boldsymbol{g}$ are defined as:

$$\boldsymbol{g}(i,i)=\frac{1}{1-4\beta_i \sin^2\left(\frac{\omega l\sqrt{LC}}{2}\right)} \tag{D.1}$$

Considering that

$$\sin^2\left(\frac{\omega\sqrt{LC}l}{2}\right)=-\left\{\frac{e^{j\omega\sqrt{LC}l}+e^{-j\omega\sqrt{LC}l}-2}{4}\right\} \tag{D.2}$$

and setting

$$z=e^{j\omega\sqrt{LC}l}=e^{j\omega\Delta t} \text{ where } \Delta t=\sqrt{LC}l=\frac{1}{2f_n}\ , \tag{D.3}$$

results in

$$\sin^2\left(\frac{\omega\sqrt{LC}l}{2}\right)=-\left\{\frac{z+z^{-1}-2}{4}\right\}. \tag{D.4}$$

Hence, $\bar{\boldsymbol{g}}(i,i)$ can be written as:

$$\bar{\boldsymbol{g}}(i,i)=\frac{1}{\beta_i}\left\{\frac{z^{-1}}{1+(1-2\beta_i)/\beta_i z^{-1}+z^{-2}}\right\}=\frac{a_i^g z^{-1}}{1+b_i^g z^{-1}+b_2^g z^{-2}} \tag{D.5}$$

Furthermore, the elements of $\boldsymbol{Y_b}$ for a lossless line can be written as:

$$\boldsymbol{Y_b}(i,i)=\frac{1}{jK\sqrt{\frac{L}{C}}\sin(\omega\sqrt{LC}l)} \tag{D.6}$$

Following a similar approach, and considering that

$$\sin\left(\omega\sqrt{LC}l\right)=\frac{e^{j\omega\sqrt{LC}l}-e^{-j\omega\sqrt{LC}l}}{2j} \tag{D.7}$$

the elements of $\overline{\boldsymbol{Y}_b}$ are represented as:

$$\overline{\boldsymbol{Y}_b}(i,j)=\begin{cases}\dfrac{2}{K}\sqrt{\dfrac{C}{L}}\dfrac{z^{-1}}{1-z^{-2}}=\dfrac{a_1^b z^{-1}}{1+b_1^b z^{-1}+b_2^b z^{-2}}, & i=j\\ -\dfrac{2}{K}\sqrt{\dfrac{C}{L}}\dfrac{z^{-1}}{1-z^{-2}}=\dfrac{a_1^b z^{-1}}{1+b_1^b z^{-1}+b_2^b z^{-2}}, & i\neq j\end{cases} \tag{D.8}$$

Again, in the case of a lossless line, the elements of $\boldsymbol{Y}''_{\boldsymbol{BB}}$ can be expressed as:

$$\boldsymbol{Y}''_{\boldsymbol{BB}}(i,i)=j\sqrt{\frac{C}{L}}\tan\left(\frac{\omega l\sqrt{LC}}{2}\right). \tag{D.9}$$

Following a similar approach, the elements of $\boldsymbol{Y}''_{\boldsymbol{BB}}$ may be written as:

$$\overline{\boldsymbol{Y}}''_{\boldsymbol{BB}}(i,i)=\sqrt{\frac{C}{L}}\left\{\frac{1-z^{-1}}{1+z^{-1}}\right\}=\frac{a_0^{BB}+a_1^{BB}z^{-1}}{1+b_1^{BB}z^{-1}},\quad a_1^{BB}=-a_0^{BB} \tag{D.10}$$

while the elements of ζ are:

$$\overline{\zeta}(i,i)=2\overline{\boldsymbol{Y}}''_{BB}(i,i)=\frac{a_0^{\zeta}+a_1^{\zeta}z^{-1}}{1+b_1^{\zeta}z^{-1}},\quad a_1^{\zeta}=-a_0^{\zeta} \tag{D.11}$$

For the case of a *lossy* line, the order of these approximations is increased by one, so that losses can be taken into account yielding:

$$g(i,i)=\frac{a_1^g z^{-1}+a_2^g z^{-2}}{1+b_1^g z^{-1}+b_2^g z^{-2}+b_3^g z^{-3}} \tag{D.12}$$

$$Y_b(i,j)=\begin{cases}\dfrac{a_1^b z^{-1}+a_2^b z^{-2}}{1+b_1^b z^{-1}+b_2^b z^{-2}+b_3^b z^{-3}}, & i=j\\ -\dfrac{a_1^b z^{-1}+a_2^b z^{-2}}{1+b_1^b z^{-1}+b_2^b z^{-2}+b_3^b z^{-3}}, & i\neq j\end{cases} \tag{D.13}$$

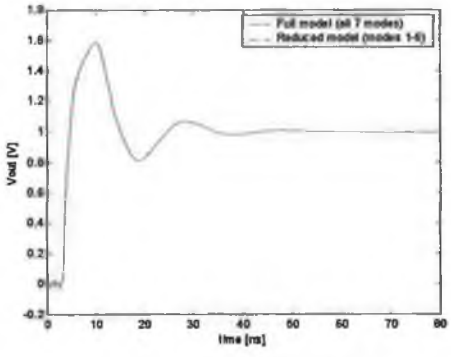
$$\zeta(i,i)=2Y''_{BB}(i,i)=\frac{a_0^{\zeta}+a_1^{\zeta}z^{-1}}{1+b_1^{\zeta}z^{-1}},\quad a_1^{\zeta}=-a_0^{\zeta} \tag{D.14}$$

APPENDIX E

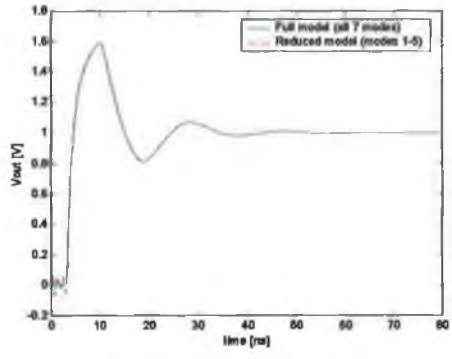
The reduced-order model responses

The responses from the various reduced-order resonant models are given in this Appendix. A full model has seven modes.

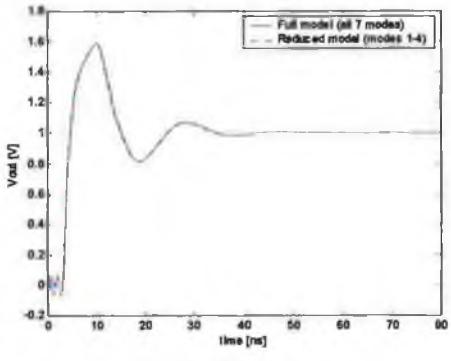
Output voltage diagrams



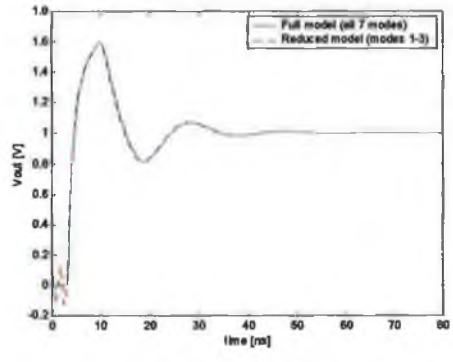
Modes (1-6) included



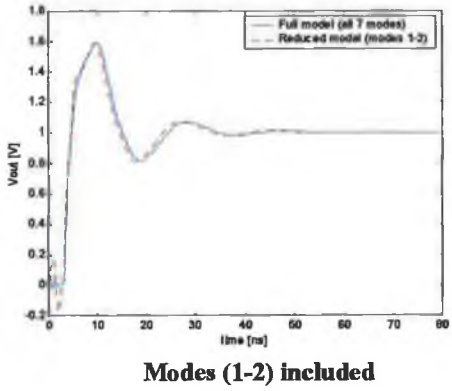
Modes (1-5) included



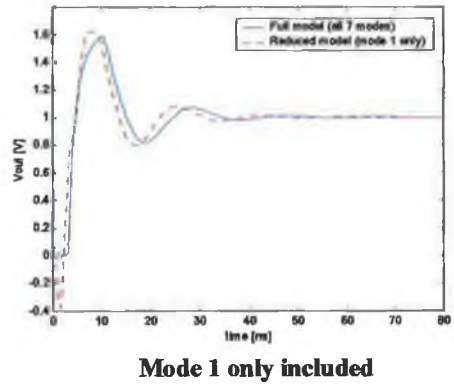
Modes (1-4) included



Modes (1-3) included

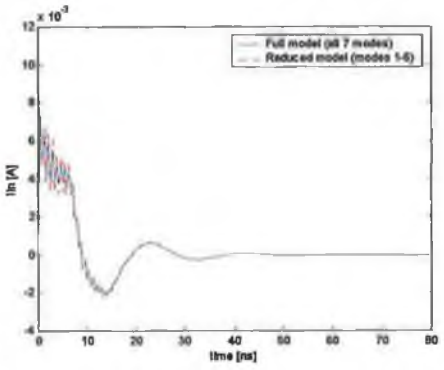


Modes (1-2) included

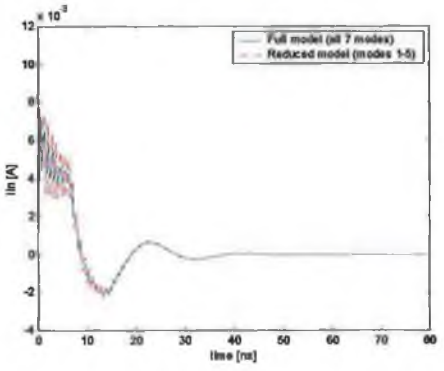


Mode 1 only included

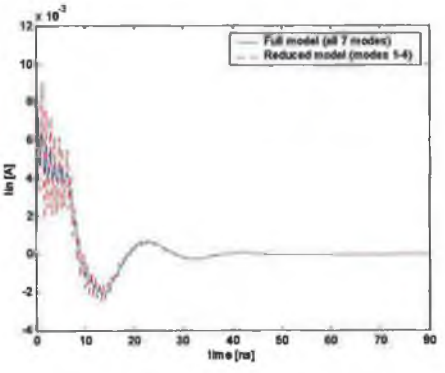
Input currents diagrams



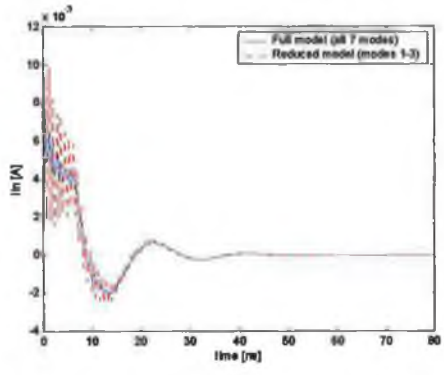
Modes (1-6) included



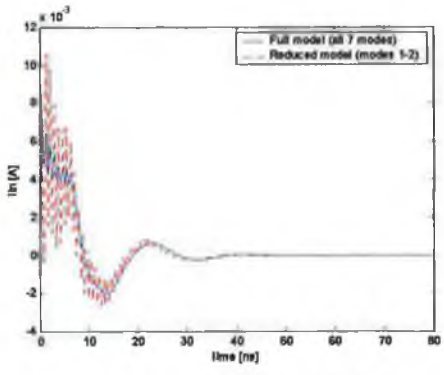
Modes (1-5) included



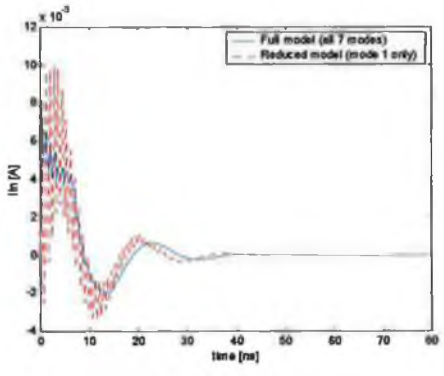
Modes (1-4) included



Modes (1-3) included



Modes (1-2) included



Mode 1 only included

APPENDIX F

Functional analysis

This Appendix gives some essential elements of functional analysis that are of interest for the wavelet-like collocation scheme presented in Chapter 7.

Metric space

A *metric space* is a set S with a global distance function (the metric g) that, for every two points x, y in S, gives the distance between them as a nonnegative real number $g(x,y)$. A metric space must also satisfy:

- $g(x,y)=0$, *iff* $x=y$
- $g(x,y)=g(y,x)$
- $g(x,y)+g(y,z)\geq g(x,z)$ (the triangle inequality)

Cauchy sequence

A sequence a_1, a_2, ... is called the *Cauchy sequence* if the metric $g(a_m,a_n)$ satisfies

$$\lim_{\min(m,n)\to\infty} g(a_m,a_n)=0.$$

Complete metric space

A *complete metric space* is a metric space in which every Cauchy sequence is convergent.

Hilbert space $L^2(\mathbb{R})$

A *Hilbert space* is a vector space H with an inner product $\langle f,g\rangle$ such that the *norm* defined by

$$|f|=\sqrt{\langle f,f\rangle}$$

turns H into a complete metric space.

An *Infinite-dimensional Hilbert space* $L^2(\mathbb{R})$ is the set of all functions $f:\mathbb{R}\to\mathbb{R}$ such that the integral of f^2 over the whole real line is finite. In this case, the *inner product* is

$$\langle f,g\rangle=\int_{-\infty}^{+\infty} f(x)g(x)dx.$$

Sobolev space $H_0^2(I)$

Let I denote a finite interval $I=[0,L]$, where L is assumed to be $L>4$.

Denote by $H^2(I)$ and $H_0^2(I)$ the following two *Sobolev spaces:*

$$H^2(I)=\left\{f(x),x\in I:\left\|f^{(i)}\right\|_2<\infty,i=0,1,2\right\}$$

$$H_0^2(I)=\left\{f(x)\in H^2(I):f(0)=f'(0)=f(L)=f'(L)=0\right\}.$$

The space $H_0^2(I)$ is a Hilbert space [CW96] equipped with the *inner product*:

$$\langle f,g\rangle=\int_I f''(x)g''(x)dx$$

and thus,

$$\|f\|=\sqrt{\langle f,f\rangle}$$

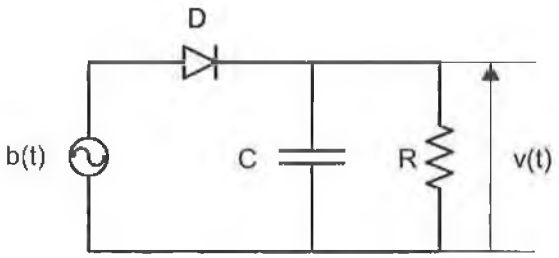
provides a *norm* for $H_0^2(I)$.

APPENDIX G

Sample systems employed in Chapter 8

In this appendix the sample systems employed in Chapter 8 are described in detail.

Diode rectifier circuit



The equation describing this diode rectifier circuit is:

$$\frac{dv(t)}{dt} = \frac{1}{C}\left[I_d \left(e^{\frac{q(b(t)-v(t))}{NkT}} - 1 \right) - \frac{v(t)}{R} \right], \qquad v(0)=0$$

or

$$\frac{dv(t)}{dt} + \frac{v(t)}{R} - \frac{I_d}{C} e^{\frac{q(b(t)-v(t))}{NkT}} = -\frac{I_d}{C}, \qquad v(0)=0$$

where

R = 100 Ω

C = 1 μF.

The current through the diode is characterized by

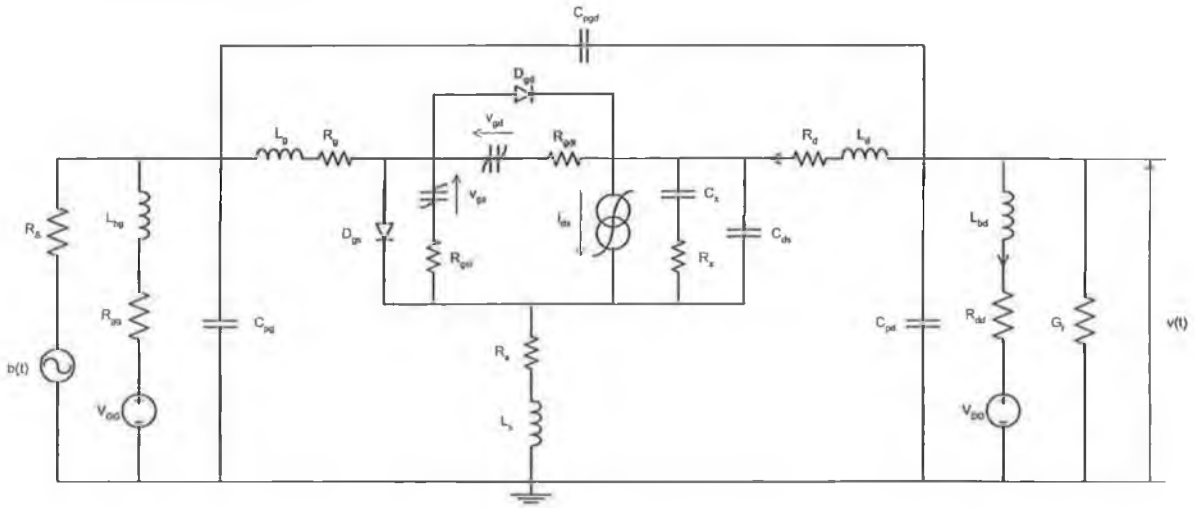
$$i_d = I_d \left[e^{\frac{q(b(t)-v(t))}{NkT}} - 1 \right]$$

term, where

I_d = 1 pA

$$\frac{q}{NkT} = 1/0.0259$$

MESFET amplifier



The equations describing the MESFET are:

$$\frac{dv_{gs}}{dt}=\frac{1}{C_{gs}}\left[\frac{v_{gd}-v_{gs}+v_{ds}+R_{gdi}\left(i_s-i_d-i_{gs}-i_{gd}\right)}{R_{gsi}+R_{gdi}}\right]=f_1\left(v_{gs},v_{gd},v_{ds},i_s,i_d\right)$$

$$\frac{dv_{gd}}{dt}=\frac{1}{C_{gd}}\left[\frac{v_{gs}-v_{gd}-v_{ds}+R_{gsi}\left(i_s-i_d-i_{gs}-i_{gd}\right)}{R_{gsi}+R_{gdi}}\right]=f_2\left(v_{gs},v_{gd},v_{ds},i_s,i_d\right)$$

$$\frac{dv_{ds}}{dt}=\frac{1}{C_{ds}}\left[i_s-\frac{v_{gd}-v_{gs}+v_{ds}+R_{gdi}\left(i_s-i_d-i_{gs}-i_{gd}\right)}{R_{gsi}+R_{gdi}}-i_{gs}-i_{ds}-\frac{v_{ds}-v_{xx}}{R_x}\right]=f_3\left(v_{gs},v_{gd},v_{ds},v_{xx},i_s,i_d\right)$$

$$\frac{dv_{xx}}{dt}=\frac{1}{C_x}\left[\frac{v_{ds}-v_{xx}}{R_x}\right]=f_4\left(v_{ds},v_{xx}\right)$$

$$\frac{di_s}{dt}=\frac{1}{L_s}\left[v_{dx}\left(v_{gs},v_{gd},v_{ds},v_g,v_d,i_s,i_d\right)-v_{ds}-i_sR_s\right]=f_5\left(v_{gs},v_{gd},v_{ds},v_g,v_d,i_s,i_d\right)$$

$$\frac{di_d}{dt}=\frac{1}{L_d}\left[v_d-i_dR_d-v_{dx}\left(v_{gs},v_{gd},v_{ds},v_g,v_d,i_s,i_d\right)\right]=f_6\left(v_{gs},v_{gd},v_{ds},v_g,v_d,i_s,i_d\right)$$

$$\frac{dv_g}{dt}=\frac{\left(C_{pgd}+C_{pd}\right)\left[\left(e_g-v_g\right)G_s-i_{lg}-i_g\right]+C_{pgd}\left[-G_lv_d-i_{ld}-i_d\right]}{C_{pg}C_{pd}+C_{pgd}\left(C_{pg}+C_{pd}\right)}=f_7\left(v_g,v_d,i_d,i_{lg},i_{ld}\right)$$

$$\frac{dv_d}{dt}=\frac{C_{pgd}\left[\left(e_g-v_g\right)G_s-i_{lg}-i_g\right]+\left(C_{pgd}+C_{pg}\right)\left[-G_lv_d-i_{ld}-i_d\right]}{C_{pg}C_{pd}+C_{pgd}\left(C_{pg}+C_{pd}\right)}=f_8\left(v_g,v_d,i_d,i_{lg},i_{ld}\right)$$

$$\frac{di_{lg}}{dt}=\frac{1}{L_{bg}}\left[v_g-R_{gg}i_{lg}-V_{GG}\right]=f_9\left(v_g,i_{lg}\right)$$

$$\frac{di_{ld}}{dt}=\frac{1}{L_{bd}}\left[v_d-R_{dd}i_{ld}-V_{DD}\right]=f_{10}\left(v_d,i_{ld}\right)$$

If the voltage drop over R_{gdi} is neglected, i.e. $R_{gdi}=0$, then

$$v_{dx}\Big|_{R_{gdi}=0}=\frac{L_gL_s\left(v_d-i_dR_d\right)+L_dL_s\left(v_g-v_{gd}-R_gi_g\right)+L_dL_g\left(R_si_s+v_{ds}\right)}{L_gL_s+L_dL_s+L_dL_g}$$

and if the voltage drop is not neglected then

$$v_{dx}\Big|_{R_{gdi}\neq 0}=v_{dx}\Big|_{R_{gdi}=0}+\frac{L_sL_dR_{gdi}\left[R_{gsi}\left(i_g-i_{gs}-i_{gd}\right)-v_{ds}-v_{gd}+v_{gs}\right]}{\left(L_gL_s+L_dL_s+L_dL_g\right)\left(R_{gdi}+R_{gsi}\right)}.$$

The currents through the diodes are given as :

$$i_{gs}=IS\_GS\left[e^{\frac{qv_{gs}}{NkT}}-1\right]$$

$$i_{gd}=IS\_GD\left[e^{\frac{q\left(v_{gs}-v_{ds}\right)}{NkT}}-1\right]$$

The formula for the drain current is:

$$v_1=\begin{cases}v_{gs}\left(1+\beta\left(V_{OUTO}-v_{ds}\right),\right. & v_{ds}\geq 0\\ v_{gd}\left(1+\beta\left(V_{OUTO}+v_{ds}\right),\right. & v_{ds}<0\end{cases}$$

$$v_{pmax}=\frac{-2A_2+\sqrt{4A_2^2-12A_1A_3}}{6A_3}$$

$$i_{dso}=\begin{cases}A_0+v_1\left(A_1+v_1\left(A_2+v_1A_3\right)\right), & v_1\geq v_{pmax}\\ A_0+v_{pmax}\left(A_1+v_{pmax}\left(A_2+v_{pmax}A_3\right)\right), & v_1<v_{pmax}\end{cases}$$

where:

$\beta = 0.05396\ \frac{1}{\mathrm{V}}$

$\gamma = 7.680735\ \frac{1}{\mathrm{V}}$

$V_{OUTO} = 19.996182\ \mathrm{V}$

$A_0 = 0.17229\ \mathrm{A}$

$A_1 = 0.093461\ \frac{\mathrm{A}}{\mathrm{V}}$

$A_2 = -0.053499\ \frac{\mathrm{A}}{\mathrm{V}^2}$

$A_3 = -0.028237\ \frac{\mathrm{A}}{\mathrm{V}^3}$

If $v_{ds}>0$ and $i_{dso}<0$, then $i_{dso}=0$. Furthermore:

$$\chi_2=-2\gamma v_{ds}$$

$$\beta_1=e^{\chi_2}$$

$$i_{ds}=i_{dso}\frac{1-\beta_1}{1+\beta_1}$$

Let $\varphi = FC \cdot VBI$. Then the capacitances C_{gs} and C_{gd} are given as:

$$C_{gs} = \begin{cases} \dfrac{CGSO}{\sqrt{1-\dfrac{v_{gs}}{VBI}}}, & v_{gs} < \varphi \\ \dfrac{CGSO\left(1-1.5FC+0.5\dfrac{v_{gs}}{VBI}\right)}{(1-FC)^{\frac{3}{2}}}, & v_{gs} \geq \varphi \end{cases}$$

and

$$C_{gd} = \begin{cases} \dfrac{CGDO}{\sqrt{1-\dfrac{v_{gd}}{VBI}}}, & v_{gd} < \varphi \\ \dfrac{CGDO\left(1-1.5FC+0.5\dfrac{v_{gd}}{VBI}\right)}{(1-FC)^{\frac{3}{2}}}, & v_{gd} \geq \varphi \end{cases}$$

The numerical parameter values are:

$G_s = 0.02$ S	$R_g = R_g + R_{gext}$	N = 1.0
$G_l = 0.02$ S	$R_d = 5.0686\ \Omega$	VBI = 1.5 V
$L_{bg} = 100$ nH	$R_{dext} = 0.25\ \Omega$	CGSO = 0.7243 pF
$L_{bd} = 100$ nH	$R_d = R_d + R_{dext}$	CGDO = 0.0197 pF
$R_{gg} = 0.0\ \Omega$	$R_s = 3.6953\ \Omega$	C_{gd} = CGDO
$R_{dd} = R_{gg}$	$R_{sext} = 0.1\ \Omega$	FC = 0.5
$C_{pg} = 0.15$ pF	$R_s = R_s + R_{sext}$	$C_x = 0.19$ pF
$C_{pd} = 0.15$ pF	$C_{pgd} = 0.01$ pF	$R_x = 800\ \Omega$
$L_g = 0.35$ nH	$R_{gdi} = 5.5\ \Omega$	$C_{ds} = 0.1037$ pF
$L_d = 0.35$ nH	$R_{gsi} = 3.18\ \Omega$	$V_{GG} = -0.2$ V
$L_s = 0.08$ nH	T = 300.15 K	$V_{DD} = 3.0$ V
$R_g = 7.21\ \Omega$	IS_GS = 1.0×10^{14} A	$\frac{kT}{q}$ @ 300K = 0.0258512607 V
$R_{gext} = 0.25\ \Omega$	IS_GD = 0.0 A	

The initial values at *t=0* are:

v_{gs} = -4.545608181458599e-001	i_d = 6.707264726002488e-002
v_{gd} = -2.843267418282757e+000	v_g = -2.000000000000000e-001
v_{ds} = 2.388706600136897e+000	v_d = 3.000000000000000e+000
v_{xx} = 2.388706600136897c+000	i_{lg} = 4.000000000010000e-003
i_s = 6.707264726001487e-002	i_{ld} = -1.270726472600249e-001

APPENDIX H

List of relevant publications

The list of publications relevant for this dissertation is given in this Appendix.

Journal papers:

[DCB05] E. Dautbegovic, M. Condon and C. Brennan, "An efficient nonlinear circuit simulation technique", in *IEEE Transactions on Microwave Theory and Techniques*, February 2005, pp. 548 – 555.

[CDB05] M. Condon, E. Dautbegovic and C. Brennan, "Efficient transient simulation of interconnect networks", to be published in *COMPEL journal*, 2005.

Conference papers:

[DCB04a] E. Dautbegovic, M. Condon and C. Brennan, "An efficient nonlinear circuit simulation technique", in *Proc. IEEE Radio Frequency Integrated Circuit Symposium (RFIC)*, Fort Worth, TX, USA, June 2004, pp. 623-626.

[DCB04b] E. Dautbegovic, M. Condon and C. Brennan, "An efficient wavelet-based nonlinear circuit simulation technique with model order reduction", in *Proc. 9th IEEE High-Frequency Postgraduate Colloquium*, Manchester, UK, September 2004, pp. 119-124.

[CD03a] M. Condon and E. Dautbegovic, "Efficient modeling of interconnects in high-speed circuits", in *Proc. European Conference on Circuit Theory and Design (ECCTD)*, Vol. 3, Kraków, Poland, September 2003, pp. 25-28.

[CD03b] M. Condon and E. Dautbegovic, "A novel envelope simulation technique for high-frequency nonlinear circuits", in *Proc. 33rd European Microwave Conference (EuMC)*, Vol. 2, Munich, Germany, October 2003, pp. 619-622.

[DC03] E. Dautbegovic and M. Condon, "Efficient simulation of interconnects in high-speed circuits", in *Proc. 8th IEEE High-Frequency Postgraduate Colloquium*, Belfast, UK, September 2003, pp. 81-84.

[CDB02] M. Condon, E. Dautbegovic and T.J. Brazil, "An efficient numerical algorithm for the transient analysis of high frequency non-linear circuits", in *Proc. 32nd European Microwave Conference (EuMC)*, Milan, Italy, September 2002.