# TRANSIENT SIMULATION OF POWER-SUPPLY NOISE IN IRREGULAR ON-CHIP POWER DISTRIBUTION NETWORKS USING LATENCY INSERTION METHOD, AND CAUSAL TRANSIENT SIMULATION OF INTERCONNECTS CHARACTERIZED BY BAND-LIMITED DATA AND TERMINATED BY ARBITRARY TERMINATIONS

A Dissertation
Presented to
The Academic Faculty

By

Subramanian N. Lalgudi

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
in
Electrical and Computer Engineering

School of Electrical and Computer Engineering
Georgia Institute of Technology
April 2008

# TRANSIENT SIMULATION OF POWER-SUPPLY NOISE IN IRREGULAR ON-CHIP POWER DISTRIBUTION NETWORKS USING LATENCY INSERTION METHOD, AND CAUSAL TRANSIENT SIMULATION OF INTERCONNECTS CHARACTERIZED BY BAND-LIMITED DATA AND TERMINATED BY ARBITRARY TERMINATIONS

Approved by:


Dr. Madhavan Swaminathan, Advisor
*Professor, School of ECE*
*Georgia Institute of Technology*

Dr. Gabriel A. Rincon-Mora
*Asso. Professor, School of ECE*
*Georgia Institute of Technology*

Dr. Jeffrey A. Davis
*Asso. Professor, School of ECE*
*Georgia Institute of Technology*

Dr. Yingjie Liu
*Asst. Professor, School of Mathematics*
*Georgia Institute of Technology*

Dr. Emmanouil M. Tentzeris
*Asso. Professor, School of ECE*
*Georgia Institute of Technology*



Date Approved: March 26, 2008

*To my family, especially to ammi jaan.*

# ACKNOWLEDGMENT

I feel proud to receive a doctoral degree. Though my effort plays a part in this accomplishment, it may not have not been possible without the people who inspired me at various stages of my life, without the oppurtunities provided by some individuals, without the help and support provided by my loving family, generous philanthropists, understanding friends, and without the insight provided by some of my knowledgeable colleagues. In my opinion, I am the small-signal gain in transistors, and you have been my DC operating point. I am indebted to you, and I thank you more than what language has allowed me to here.

I would like to thank my Ph.D. advisor, Prof. Madhavan Swaminathan, for giving me an oppurtunity to study in Georgia Tech and be part of his group. Being in his group, I was exposed to various aspects of signal and power integrity, which I, otherwise, can only get exposed to in semiconductor-based product and CAD companies. Naturally, coming out of his group made us a darling of these companies. I would like to thank him for making me work on important problems.

I would like to thank my M.S. degree advisor, Prof. Shanker Balasubramaniam, in Iowa state university. Working with him and in computational electromagnetics had a significant influence on my learning and on my work afterwards. His approach to understanding and reducing complex problems (e.g., volume/surface equivalence theorems in electromagnetics) to related and known simpler ones (Thevenin's theorem in circuits) stuck with me.

I would like to thank my Ph.D. committee members, Prof. Jeff Davis, Prof. Manos Tentzeris, Prof. Gabriel Rincon-Mora, and Prof. Yingjie Liu for agreeing to be part of my committee. I enjoyed the discussions we had during and after both my proposal and defense presentations, and I appreciate your suggestions.

I would like to thank Dr. Giorgio Casinovi for providing me with insights to

developing better convolution formulations and to analytically verifying numerical convolution results. I also would like to thank Dr. Ege Engin for his help in my research and Dr. Mahadevan Iyer for useful discussions.

I would like to thank the following colleagues for their help and cooperation: Jifeng Mao, Bhyrav Mutnury, Amit Bavisi, Souvik Mukherjee, Rohan Mandrekar, Vinu Govind, Sungwan Min, Tae Hong Kim, Krishna Srinivasan, Prathap Muthana, Wansuk Yun, Krishna Bharat, Nevin Altunyurt, Abdemanaf Tambawala, Marie-Solange Milleron, Janani Chandresekhar, Abhilash Goyal, Kijin Han, Bernard Yang, Narayanan, Susan, Eddy, Myungyun, Sukruth, Vishal, Aswini, Ranjith, Rishiraj, Nithya, and Tapobrata. I would like to thank my friends Raj, Ganesh, Vishwa, Maryam, Vidya, Moumita, Vijay, and Guru for their help and support.

I owe my sincere gratitude to philanthropists and organizations who helped with my education in my high-school and undergraduate years. Special mention to Mr. Krishnamoorthy (Binny Corporation, Chennai), Mr. Raja (R.M.H Corporation, Chennai), Lakshmi Charities (TVS Co., Chennai), Ms. Elizabeth Thomas (Correspondent, St. Marys, Chennai), three more persons whose names I do not know, Sumati Vishal and Rajasthan Youth Association. Special mention to Mr. Rangarajan (Teacher), Mr. Hari, and Mr. Subramaniam (Teacher, St. Marys) for referring me to some of the above philanthropists.

I am fortunate and feel proud to have been born to my parents and into my family. My father, Mr. Natarajan, has been my advisor, and my elder brother, Venkataraman, has been a symbol of hardwork. I thank my younger brother, Sabdharishi, for sharing most of my domestic responsibilities during my Ph.D. degree. My elder sister, Bhuvana, for being a mom to us when mother, Ms. Saraswathi, was away working.

Mom, I sincerely appreciate your efforts, which in my opinion are extraordinary, in putting me through good schools in spite of the obstacles you had to face. I remember the struggles you had to face, the hardwork you put in, and the determination you

showed. I have imbibed the successful values I learnt seeing you work: the importance of having pride in one's work, and being *numero uno* in what you do. You've been the pillar I've leant against. *You are my hero, ammi jaan!* Hopefully, I can make you feel prouder in the future.

# TABLE OF CONTENTS

## Part I : Time-Domain Simulation of Power-Supply Noise in On-Chip Power Distribution Networks Using a FDTD-like Method

# LIST OF TABLES

# LIST OF FIGURES

# SUMMARY

Power distribution networks (PDNs) are conducting structures employed in semiconductor systems with the aim of providing circuits with reliable and constant operating voltage. This network has non-neglible electrical parasitics. Consequently, when digital circuits inside the chip switch, the supply voltage delivered to them does not remain ideal and exhibits spatial and temporal voltage fluctuations. These fluctuations in the supply voltage, known as the power-supply noise (PSN), can affect the functionality and the performance of modern microprocessors. The design of this PDN in the chip is an important part in ensuring power integrity. Modeling and simulation of the PSN in on-chip PDNs is important to reduce the cost of processors. These PDNs have irregular geometries, which affect the PSN. As a result, they have to be modeled. The problem sizes encountered in this simulation are usually large (on the order of millions), necessitating computationally efficient simulation approaches. Existing approaches for this simulation do not guarantee at least one of the following three required properties: computationally efficiency, accuracy, and numerically robustness. Therefore, there is a need to develop accurate, numerically robust, and efficient algorithms for this simulation.

For many interconnects (e.g., transmission lines, board connectors, package PDNs), only their frequency responses and SPICE circuits (e.g., nonlinear switching drivers, equivalent circuits of interconnects) terminating them are known. These frequency responses are usually available only up to a certain maximum frequency. Simulating the electrical behavior of these systems is important for the reliable design of microprocessors and for their faster time-to-market. Because terminations can be nonlinear, a transient simulation is required. There is a need for a transient simulation of band-limited frequency-domain data characterizing a multiport passive system with SPICE circuits. The number of ports can be large ($\geq$ 100 ports). In this simulation, unlike

in traditional circuit simulators, normal properties like stability and causality of transient results are not automatically met and have to be ensured. Existing techniques for this simulation do not guarantee at least one of the following three required properties: computationally efficiency for a large number of ports, causality, and accuracy. Therefore, there is a need to develop accurate and efficient time-domain techniques for this simulation that also ensure causality.

The objectives of this Ph.D. research are twofold: 1) To develop accurate, numerically robust, and computationally efficient time-domain algorithms to compute PSN in on-chip PDNs with irregular geometries. 2) To develop accurate and computationally efficient time-domain algorithms for the causal cosimulation of band-limited frequency-domain data with SPICE circuits.

# CHAPTER 1

# INTRODUCTION

This dissertation consists of two parts. The first part is about developing accurate, numerically robust, and computationally efficient time-domain-based numerical techniques for simulating power-supply noise (PSN) in on-chip power distribution networks (PDNs). The second part is about developing accurate and computationally efficient time-domain numerical techniques for causal transient simulation of interconnects characterized by band-limited (b.l.) frequency-domain (f.d.) data and terminated by arbitrary SPICE (standing for Simulation Program with Integrated Circuit Emphasis) [3] circuits. The techniques developed in the second part are expected to be useful for simulating PSN in on-chip PDNs (SPICE circuits) including the effect of the package PDNs (frequency-domain data).

The rest of this chapter is organized as follows: In Section 1.1, the need for the problem focussed by the first part of this dissertation is justified. In Section 1.2, the history of the research related to this part is described. The same procedure is repeated for the problem focussed by the second part of this dissertation in Sections 1.3 and 1.4. Based on these histories, the research proposed in this dissertation is described in Sections 1.2 and 1.4 and stated in Section 1.5. In Section 1.6, the completed research is discussed. Finally, in Section 1.7, the outline of the rest of this dissertation is provided.

## 1.1 Need for Simulating PSN in On-Chip PDNs

### Origin of Power-Supply Noise in Microprocessors

Power is supplied to digital circuits in chip (see Figure 1) from the voltage regulator module (VRM) in the printed circuit board (PCB) through stages of conducting structures collectively known as the power distribution network (PDN) (or the power supply network). Conducting structures have inductance. Since the VRM is far away

from the switching circuits, this inductance can be large. This network is usually made up of copper. Consequently, the PDN also has nonzero resistance. The nonzero

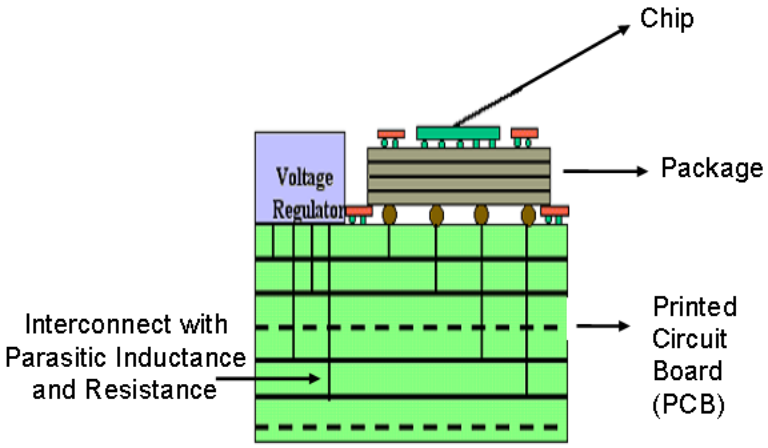



**Figure 1. Simplified 3-D view of the power distribution network of a modern digital system.**

parasitics of the PDN creates one fundamental problem: *the supply voltage observed at the terminals of switching circuits is not same as the voltage that was supplied by the VRM*. Specifically, because of the PDN's nonzero resistance, the supply voltage across the circuit's supply terminals is less than what is supplied by the VRM. The magnitude of this difference depends directly on the resistance of the PDN. Because of the PDN's nonzero inductance, the supply voltage seen at a circuit fluctuates with time (see Figure 2) every time this circuit or any other circuit in the chip switches. This temporal voltage fluctuation is a direct consequence of Faraday's law. By this law, voltage is induced in a conductor subjected to a time-varying magnetic field. The magnitude of the temporal voltage fluctuation depends directly on the total (loop) inductance of the PDN, how fast the circuits switch, and how many of them switch at the same time. The difference between the voltage supplied by the VRM and the voltage actually received by the circuits is referred to as the *power-supply noise* (PSN) [4], [5], [6], [7]. This noise is also known as the *switching noise* or the *simultaneous switching noise* (SSN).

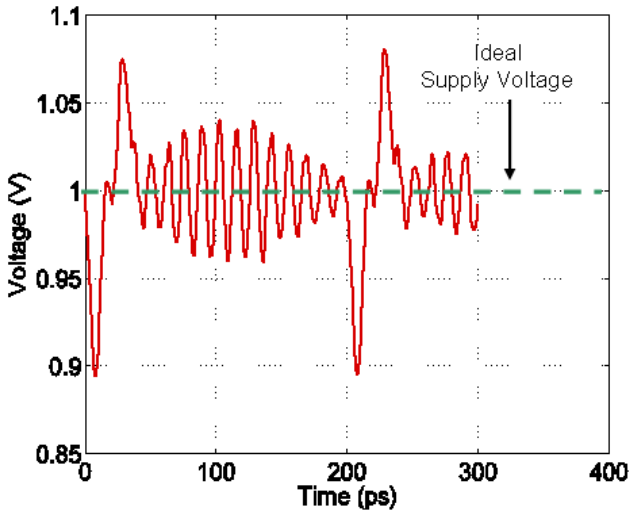



**Figure 2. Temporal fluctuation of the voltage observed between the power and ground terminals of an on-chip circuit.**

### Undesirable Effects of PSN on Functionality and Performance

The PSN can either degrade the performance of the processor (see [4]) by slowing the processor down [8], [9] and/or by affecting the timing of the clock signals [10] [11] or affect the functionality of the processor by causing logic failures [12] in switching circuits. It is estimated that a 10% fluctuation in the supply voltage may translate to more than a 10% timing uncertainty. Since the PDN is always going to be nonideal, PSN can not be eliminated completely – it can only be controlled. Fortunately, the performance of the processor is not affected much, if the magnitude of PSN is kept within a small fraction of the ideal supply voltage. Usually, this fraction is less than 0.1 [13].

### Worsening PSN Trends with Device Scaling

Unfortunately, ensuring that the PSN stays within 10% of the actual supply voltage is becoming tougher with the scaling of transistors [2], [12], [13], [14], [15], [16]: Device sizes have shrunk, while chip sizes have not. This means that there are going to be more circuits that are going to switch simultaneously. As the device sizes have shrunk

to accomodate faster transistors, the switching speed is going to more with scaling. To maintain a constant electric field in transistor gates, the supply voltage is reduced. This reduction in supply voltage means that the absolute magnitude of PSN has to get smaller with scaling. It is described in [14] that with a constant-field scaling, the signal-to-noise ratio as a result of $\frac{di}{dt}$ noise scales as $1/S^4$ when scaling the process by $1/S$. With scaling, the total current needed increases. As the interconnect resistance increases with scaling, the IR drop increases with scaling. It has been reported in [17] that based on a survey of over 206 tapeouts, targeting process technology of 0.13 micron or greater, more than 50% of tapeouts will fail if the power distribution system is not validated beforehand.

Moreover, with reduced supply voltage, transistor leakage current [18] (subthreshold, gate leakage, etc.) increases [18], [19], [20]. This current causes IR drop [20], [21]. This drop eats into the small margin allowed for the PSN. With the prediction that the leakage power dissipation is going to be more than the switching power dissipation for 65 nm node and below [19], the IR drop from the leakage current can become serious.

**Decoupling Capacitors for Effectively Controlling PSN**

The most common strategy to control PSN is to reduce the total loop impedance between switching circuits and the VRM [22]. Other ways, like reducing the number of circuits that switch simultaneously, affect the performance of the processor and therefore are not preferred [22]. A simple strategy to reduce the effective impedance is to size the structures: 1) Increasing the width of the line reduces the line's resistance. 2) Reducing the spacing between lines reduces the total loop inductance. 3) Increasing the number of metal layers reduces the effective impedance of lines. However, this strategy alone is not sufficient for reducing the effective impedance. The total loop inductance is naturally reduced if the VRM is brought closer to switching circuits. As the VRM is bulky and cannot be integrated into the chip, the loop inductance

has to be reduced through other ways. In practice, this reduction is accomplished by placing capacitors between power and ground terminals at points close to switching circuits [23], [24], [25], [26]. These capacitors act as local power supplies (i.e., local VRMs) that give the charge necessary for switching circuits. As they are closer to switching circuits than they are to the VRM, the loop inductance is reduced. They get recharged completely (from the VRM) when circuits stop switching. These capacitors are known as *decoupling capacitors* (*decaps* in short).

In today's processors, almost 10% of the chip area is allocated for decoupling capacitors [13]. Considering the large number of processors built nowadays, each decoupling capacitor added increases the cost of the individual processor significantly. *Therefore, to reduce the overall cost, the total decoupling capacitance has to be kept small, just small enough to ensure reliable performance.* Placing all decoupling capacitors near switching circuits is not possible, as this process requires a significant chip real estate that is not affordable. The usual practice is to place decoupling capacitor at every level of the processor (see Figure 3) [26]: i.e., at the VRM level, at the board level, at the package level, and at the chip level. This way the total capacitance

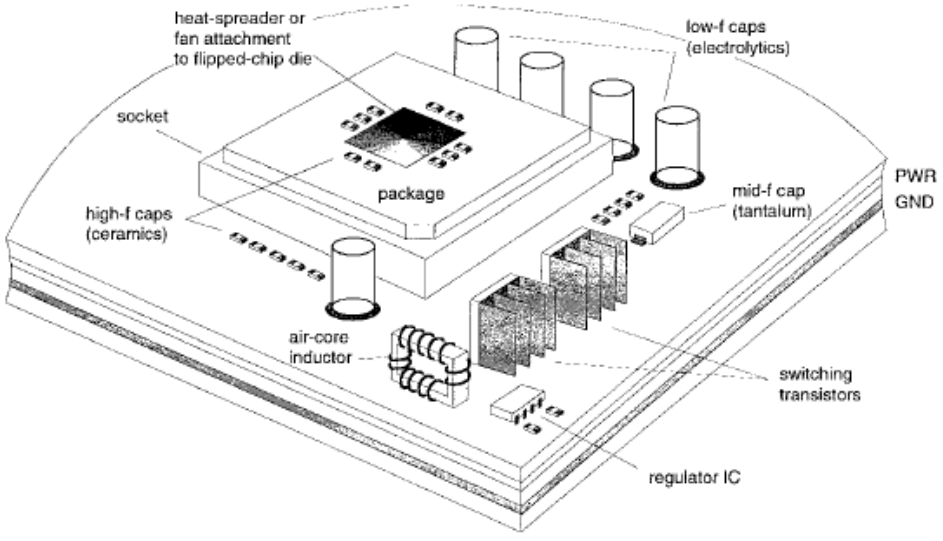



**Figure 3. Decoupling capacitor at different levels of the PDN. Source: [1]**

required in the chip is kept small, requiring anywhere between 10% to 20% of chip's total area. On-chip decoupling capacitors, unfortunately, can only provide charges to switching circuits that are near by them. However, it is difficult to determine apriori which part of the chip will switch, how many circuits will switch simultaneously, and when they will switch. So before processors are mass produced, the PSN problem has to addressed. Considering the lack of the switching information mentioned above, this problem is addressed assuming worst-case conditions.

**Need for Modeling and Simulation of PSN**

The natural strategy to build the prototype first and then redesign it for the PSN problem is not cost effective. The PSN problem is hardly addressed this way. Modeling and simulating the PSN (to ensure that the PSN is within limits) before prototypes are built is the most common and cost-effective way. When this simulation is performed just before the prototypes are built, this simulation is referred to as the *post-layout* simulation. At this design stage, all the information of the processor, like the geometry of the PDN, the location of active circuits, are known. This simulation involves computing the PSN given the geometry and switching sources information. Accuracy is the most important criterion for this simulation. Most accurate models for PDNs and switching circuits are employed, as simulation results have to correlate with measurement results from the prototype. *Consequently, post-layout simulation consumes a lot of time and memory. Almost all ocommercial computer-aided design (CAD) vendors develop algorithms for this type of simulation.*

**Need for CAD tools for PSN analysis in Pre-layout Stage**

Redesign of prototype may be necessary. When this is the only time the PSN is simulated, the redesign effort can be significant, sometimes requiring a complete redesign of the processor. To ensure the redesign effort at the post-layout simulation is minimum, PSN has to be simulated even before the final layout is ready. The prescribed

practice is to simulate PSN at every stage of the design [27], starting from the stage where the design is first conceived. This simulation is considerably different from the simulation at the post-layout stage. Unlike the post-layout stage, not everything is known about the design, and designs are more likely and frequently to be changed, not necessarily as part of the PSN fix. Such lack of information at this stage makes the returns of the most accurate but costly solution small, as designs may be changed. Therefore, almost all post-layout simulation tools (and the approaches they advocate) are not viable for the pre-layout simulation. The pre-layout simulation tools should necessarily be more time and memory efficient than the post-layout simulation tools. As returns for the most accurate but costly solution is going to be less, only simplified models are employed in this simulation. There has been a constant research effort to balance computational complexity with the accuracy of the simulation.

**Modeling and Simulation of Package and Board PDNs**

Modeling and simulation of PSN requires a distributed modeling, as PDNs are electrically large. However, even at the pre-layout stage (so that simplified models can be used), modeling the PDN at all levels (board, package, and chip) of the processor at once and with equal detail is still out of reach in today's computers. The usual strategy is to model the PDN at one level with enough detail and assume coarse models for the rest of the PDN. Since the impedance of the PDN is highly frequency dependent [24], [26], simulation following this strategy ensures that the PSN is within limits only in a limited frequency band. Of course, the PSN has to be within limits at all frequencies [25], [26]. Early (early 1990's) research effort on PSN simulation has been on simulating package and board PDNs (see Figure 4) [28], [29], assuming simplified models for the chip and VRM. This simulation focusses on simulating the PSN behavior for frequencies usually below the chip-package resonance frequency, which is usually a few hundred megahertz. All these tools work in the frequency domain.

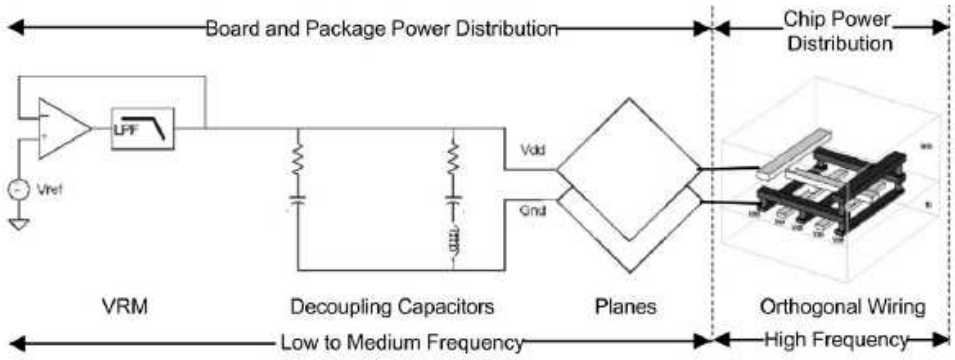



**Figure 4. Board, package, and chip power distribution network. Source: [2]**

### Need to Model and Simulate PSN in On-Chip PDNs

*The need to model and simulate PSN in on-chip PDNs is necessary for three reasons*:

1. To ensure PSN is within limits for high frequencies ($> 100$ MHz), package PDN simulation alone is not sufficient; the on-chip PDN (see Figure 4) also has to be modeled and simulated. At these frequencies, the on-chip decoupling capacitors provide a smaller impedance path than the path to the VRM.

2. Among all resonances in the input impedance observed at a point in the chip, the chip-package resonance has the maximum amplitude [24]. It is important to keep this amplitude small so that the PSN is within a 5-10% of the supply voltage of the VRM. This amplitude is not a constant and varies among different locations in the chip. Therefore, to ensure that the maximum value among these amplitudes is within tolerable limits, the on-chip PDN has to be modeled.

3. The DC (i.e., when circuits do not switch at all) supply voltages, which are part of PSN, the switching circuits receive are significantly controlled by on-chip PDNs. This control has to with the higher resistive loss of on-chip PDNs compared to that of PDNs at other levels. This high loss is because of small dimensions (width and thickness) of interconnects observed in chips. Without modeling the on-chip PDN, the budgeting for DC drops cannot be reliably completed. As a result, to ensure that DC supply voltages at all switching

circuit terminals are within limits, an on-chip power grid simulation is necessary.

**On-Chip PDN Simulation Vs. Package/Board PDN Simulation**

Simulating the PSN in on-chip PDN is different from simulating the PSN in package PDNs: 1) The on-chip PDN is a grid, while the package PDN is a plane. The PSN propagation in grid-like geometries is still not well understood. However, with a plane, analytical result is at least possible for solid planes (which can be thought of as parallel-plate waveguide). Analytical modeling of power grids is still not well developed, even for simple on-chip PDN geometries. Therefore, a numerical simulation is necessary for computing PSN in on-chip PDNs. 2) On-chip PDN tools are time-domain tools, whereas package PDN tools are usually frequency-domain tools. This difference is necessitated by the fact that switching circuits inside the chip are inherently nonlinear. Moreover, as on-chip PDNs are lossy, a time-domain tool is preferred. Time-domain numerical solution is confronted by new problems that are not normally encountered in frequency-domain solution, stability of the numerical solution, for instance. 3) The problem size encountered in a typical power grid simulation (usually in millions of nodes) is lot higher than what is usually encountered in power plane simulation. *The focus of this dissertation is on the pre-layout-level simulation of PSN in on-chip PDNs.*

## 1.2 Simulation of PSN in On-Chip PDNs

**Geometry of on-chip PDN**

The on-chip PDN is a multilayered grid arrangement inside the chip [30], [31]. An example of the on-chip PDN is shown in Figure 5. Such a PDN geometry is employed for high-performance microprocessors. In Figure 5, the PDN is immersed in silicon dioxide and is housed on top of a lossy silicon substrate. The on-chip PDN is connected to the package PDN through (power and ground) controlled-collapse-chip connection (C4) bumps at the on-chip metal layer closest to the package. All lines in

a layer are routed parallel to each other and are routed perpendicular to lines in the adjacent layers. Power (Ground) lines in adjacent layers are connected through vias at the intersection points of lines. The CMOS circuits reside inside the silicon substrate. The power-supply (ground-supply) terminal of a MOS transistor is connected to the nearest power (ground) line in the metal layer (M1) closest to the substrate.

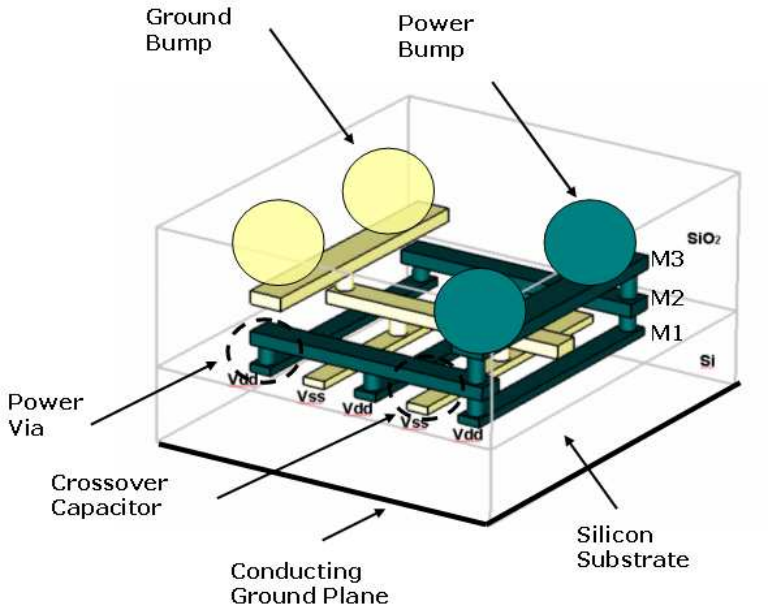



**Figure 5. A simplified model of the on-chip PDN in an high-performance microprocessor.**

### Overall Objectives of any On-Chip PDN Simulator

*The objective of the power grid simulation is to compute the spatial and temporal fluctuation in the supply voltage in the on-chip PDN. The output of the simulation are 1) the spatial voltage profile (please see Figure 29 in Chapter 4, Figure 32 in Chapter 5) in the PDN given the average currents consumed by switching circuits and/or 2) the temporal voltage profile in the PDN given the switching current profile. In both cases, the geometry of the PDN, profiles of C4 bumps, rest of the PDN, and on-chip decaps are also provided as inputs.*

**Typical Steps in On-Chip PDN Simulation**

On-chip power grid simulation involves three steps:

1. Fix the equivalent circuit for different parts of the PDN, and extract the parasitic inductance, resistance, and capacitance of the different parts.

2. Construct a distributed equivalent circuit of the PDN. The values of circuit elements are obtained from the results of Step 1.

3. Solve the resulting circuit problem for spatial and temporal voltage profiles.

**Need for Computationally Efficient Circuit Simulation**

*Among the above three steps, the third step is the most important in a pre-layout simulation.* In a pre-layout stage, the first two steps are accomplished by constructing a transmission line type distributed equivalent circuit for power and ground lines. In doing so, only adjacent capacitive coupling and neighboring inductive coupling are considered. By doing so, computational complexity is made manageable, though with a little compromise in accuracy. This simplification is one of the key differences from a post-layout simulation. Though the last step is only a circuit simulation step and SPICE is a well-known circuit simulation tool, majority of the prior work [27], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42], [43] in the on-chip power grid simulation focusses on this circuit simulation step. *The focus of this thesis is also on this step.*

*The overwhelming attention the third step has been receiving has to do with the computational difficulties faced in simulating circuits with large (> one million nodes) problem sizes.* The large problem sizes are the consequence of modeling the on-chip PDN in a distributed manner, a manner which has been discussed earlier to be necessary. *There is also another reason (a more important one) that significantly compounds the problem size: Discretization of an on-chip PDN has to be much finer than what is needed for a distributed modeling.* This new reason can be explained as

follows. The length of the smallest segment in the discretized problem is limited by the minimum spacing between a power line and the ground line nearest to it. This spacing is usually on the order of micrometers or less. This spacing is usually much smaller than one-tenth of the minimum wavelength that needs to be resolved in the excitation. The minimum wavelength that needs to be resolved in silicon dioxide assuming a 10 ps rise time (worst case) in switching current is approximately 750 um. Though transistor sizes are continuously shrunk, the chip sizes are not; chip sizes are on the order of square millimeters. Then, the number of nodes, $N_n$, in the discretized problem could be on the order of millions.

This large problem size would not matter if SPICE is efficient for such large problems. SPICE is a well-known tool for circuit simulation. The results from SPICE are usually the most accurate for a circuit simulation tool. Moreover, SPICE is robust: it can take almost all circuit element values and provide stable transient results. SPICE is also the first choice tool for small ($\leq$ 10K nodes) problem sizes. However, SPICE has been found to be computationally prohibitive for problem sizes on the order of millions [43]. Thus, *there is a clear need for an alternative tool that is computationally more efficient than SPICE for the power grid simulation (and for large problems, in general).* Considering that the circuit matrices are going to be sparse, optimal complexities for memory and run time are preferred. In other words, if $N_n$ refers to the number of nodes in the circuit, the memory and time complexities of the circuit simulation step should preferably be $O(N_n)$ each.

**Ability to Handle Irregular PDNs**

Apart from efficiency, some other features are also preferred in the alternative tool. For example, the capability to handle irregularities in the power grid. The power grids can get highly irregular during the course of the design. A power grid is referred to as being irregular when at least one of the following happens (see Figure 6): 1) When there is a nonuniform spacing between power and ground lines. 2) When a power

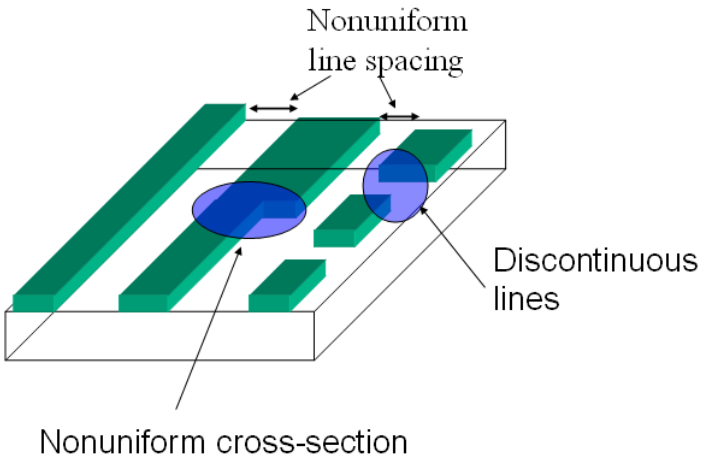



**Figure 6. Various types of irregularities in the on-chip PDN.**

(ground) line has a nonuniform cross-section along its length. This situation happens when the width of a part of the line is changed. Wire sizing is one of the first things done to control PSN in noise-sensitive areas of the chip. 3) When lines do not run continuously from one side of the chip to the other, i.e., when lines are discontinuous. These irregularities affect the total impedance seen at a given location in the circuit and therefore have to be modeled. Naturally, the parasitic extraction step needs to take these irregularities into account. It is also important for the circuit simulation step to take equivalent circuits of such irregular PDNs and simulate PSN in them. Though SPICE and most of the other SPICE-based solvers do not have a problem in this regard, it can be a problem in some other tools (which is one of the focus points of this dissertation). Finally, to improve the accuracy of the PSN computation, the accuracy needs to be ensured for all steps of the simulation, particularly the step concerning the equivalent circuit.

### 1.2.1 Specific Objectives of this Dissertation

*The specific objective of this dissertation is to develop an on-chip power grid simulation tool with the following features*:

1. The circuit simulation algorithm should

(a) have a SPICE-like accuracy,

(b) be as robust as SPICE (this means that the new algorithm should have convergence properties similar to that of SPICE),

(c) should have memory and time complexities that scale as $O(N_n)$ each, and

(d) should be able to simulate irregular power grids.

2. The tool should have accurate equivalent circuits for the PDN.

### 1.2.2 History of Prior Work and Its Limitations

As mentioned earlier, most of the prior work focus on obtaining a computationally more efficient algorithm than SPICE for the circuit simulation step. Considering the large problem sizes encountered in the on-chip power grid simulation, only a linear circuit simulation has thus far been the main focus of the community. The circuit simulation step for power grids involves the following steps:

1. Identify and construct equations describing voltages and/or currents from the equivalent circuits of PDN and switching sources.

2. Cast these equations in a suitable form, resulting in a linear system of equations.

3. Discretize the equations using a suitable numerical integration rule.

4. Solve the discretized matrix system using a suitable solver. From the solution of this system, the spatial and temporal voltage fluctuations of nodes can be computed.

All the properties of the algorithm depend on what equations are employed, on the form used to cast these equations, on what numerical integration rules are used for discretization, and, finally and *most importantly, on what type of solver is used for solving the matrix system.* Majority of the prior approaches for the power grid simulation differ from SPICE in the last step.

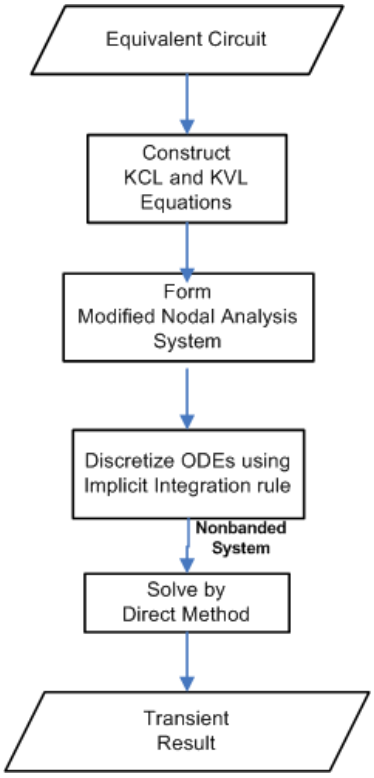



**Figure 7. Steps in SPICE for a transient simulation.**

*Computational Inefficiency of SPICE Elaborated*

The steps involved in SPICE for a transient simulation can be described as shown in Figure 7. SPICE constructs equations describing Kirchoff's current and voltage laws (KCLs and KVLs) from the (power grid) equivalent circuit given. These equations are then cast either using a modified nodal analysis (MNA) formulation. The linear system of equations are discretized using an implicit integration rule. Examples of these rules are the backward Euler and the Trapezoidal integration rules [44]. Such a rule usually results in a nonbanded linear system. The accuracy of the formulation depends on the accuracy of the integration rule. The backward Euler scheme results in a first-order accurate solution, i.e., the accuracy scales as $O((\Delta t))$, while the trapezoidal scheme results in a second-order accurate solution, i.e., the accuracy scales as $O\left((\Delta t)^2\right)$. The transient simulation using an implicit integration rule is usually unconditionally stable. Such a stability is manifested in terms of the factors deciding

the time step of the simulation. Usually, using such a rule, the time step can be any real number, less than of course the total simulation time. The choice usually is a number smaller than the smallest rise time in the excitation. SPICE uses a direct solver based on Gauss elimination to solve the discretized system. The solution from the direct solver are the most accurate. The solver does not create any convergence problems during the transient simulation. Because of the implicit integration rules, irregular geometries do not pose any simulation challenges. *The main disadvantage of using a direct solver with a nonbanded matrix is that the computational complexity depends on the way the nodes in the circuit are numbered.* For a completely random node numbering, the memory complexity can scale as $O\left(N_n^2\right)$ and time as $O\left(N_n^3\right)$. The computational complexities depending on the second and third powers of $N_n$ are not practically feasible.

#### *1.2.2.1 SPICE-based Approaches*

*Majority of the prior approaches [30], [35], [36], [38], [34], [37] replace the direct solver in the above scheme of things (see Figure 7) with a different solver.* Sometimes, a nodal analysis system is constructed instead of the modified nodal system in Figure 7. The choices for the new solver are 1) the iterative solvers [35], [30], [36], [38] and 2) the statistical solvers [34], [37]. Unlike the direct solver, the computational complexity of an iterative solver does not depend on the way nodes are numbered. Using these solvers, computationally efficiency is improved. *However, these new solvers lack either the accuracy of or the convergence property of the direct solver.*

Some of the approaches [32], [33], [43] retain the direct solver, but manage complexity by dividing the full problem into many smaller ones and solving the smaller problems. For example, in [32], an hierachical simulation is proposed. It does so by 1) partitioning the power grid equivalent circuit into subcircuits, 2) solving individual partitions using direct solver, and 3) obtaining the solution of the full circuit from

the solution of the partitions. This approach compromises some accuracy in dividing the circuit into partitions. Also, the computational complexity depends on the number of partitions and on how large each one of them is. In [33], a "shell"-based partitioning technique is proposed that specifically makes use of the locality effect in flip-chip technologies. However, the accuracy and robustness (i.e., its applicability to irregular grids, DC + transient simulation) of this technique are still not known.

In [43], direct solver is once again retained; however, the direct solver is applied to a series of tridiagonal (banded) system. Such a system can be solved using a direct solver in linear (w.r. to $N_n$) runtime and using linear memory resources. Such a system is obtained by 1) starting with the scalar wave equations in terms of unknown node voltages, 2) discretizing these (continuous) equations using suitable implicit integration rule, and 3) applying an alternate direct-implicit (ADI) algorithm [43] on these equations. However, this approach results in three problems when applied to power-grid equivalent circuits: 1) The transient simulation becomes conditionally stable depending on the boundary conditions for the voltages. 2) The transient simulation's stability can not be proven for irregular power grids. 3) Finally, ADI-inspired transient simulation requires that two orthogonal directions of line routing be present in every layer. *However, in high-performance power grids, power-ground lines in a layer are routed in only one direction. This dissertation focusses on such high-performance grids.* Therefore, the ADI algorithm [43] can not be applied for all PDN geometries. Moreover, it has been shown in this dissertation that this algorithm loses it unconditional stability when applying an open-circuit at the PDN's boundary.

#### *1.2.2.2 Finite-Difference Time-Domain (FDTD) Method Based Approaches*

*One interesting and not a well-studied approach for the transient simulation of power grids is based on applying a finite-difference time-domain-like (FDTD-like) method for circuit simulation [31], [39], [40], [45], [41]. This approach can be thought to have the steps as shown in Figure 8. The main advantage of such an approach is that it usually*

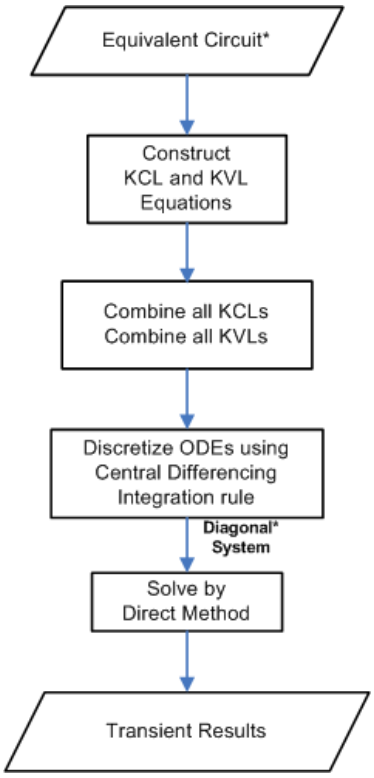



**Figure 8. Steps in FDTD-based approaches for a transient circuit simulation. The symbol '*' in the equivalent circuit denotes that this approach applies only for restricted circuits. The symbol '*' above the diagonal system denotes that this approach can result in diagonal system for some equivalent circuits.**

*results in a diagonal matrix. So direct solvers can be employed in an optimal manner.* The approaches [39], [40], [45], [41], [31] propose a solution whose complexity scales linearly with the problem size. These approaches also guarantee SPICE-like accuracy and numerical robustness. A comparison is shown in Table 1 of different simulation features between SPICE-based approaches and FDTD-based approaches for on-chip power grid simulation. *Because of SPICE-like accuracy and robustness of an FDTD-like method and the possibility of achieving a linear computational complexity with it, (these are also the objectives), this dissertation is based on an FDTD-based approach.* The objective of this dissertation is to enable an FDTD-based approach for irregular on-chip power grids.

This new approach [39], [40], [41], [31] has a lot of similarities with the FDTD method [46], which is well understood. The FDTD method for Maxwell's equations has been described in Appendix A.

*FDTD-like method for Solving Circuit Problems*

An FDTD-like method for circuits can be interpreted as a scalar version of the FDTD method. The steps in the transient circuit simulation using FDTD-based approaches is described in Figure 8. KCL and KVL equations are used in the place of Maxwell's curl equations involving magnetic fields and electric fields. The voltages and currents are the unknowns. There is a key difference between the circuit problem and the wave problem: In a circuit problem, there is not always a nonzero propagation delay between two nodes in the circuit. For example, presence of coupling capacitance between two nodes creates zero propagation delay between the nodes. Such a capacitance (as will be discussed later) can make the FDTD-like method lose its linear computational complexity per each step of the transient simulation. An FDTD-like method for circuits inherits all the merits of the original FDTD method only in a few type of circuits and inherits the demerits, unfortunately, in all the circuits.

*First Application of an FDTD-like method for Circuit Problems*

**Table 1. Comparison of different simulation features between SPICE-based approaches and FDTD-based approaches for on-chip power grid simulation**

| **Feature** | **SPICE-based Approaches** | **FDTD-based Approaches** |
|---|---|---|
| Computational Efficiency | Better than SPICE | Better than SPICE |
| Accuracy | Worse than SPICE | Same as SPICE |
| Convergence/Robustness | Can be an issue | Not an issue |
| Stability | Unconditional | Conditional |
| Formulation | Exist for all circuits | Exist only for certain circuits |
| | Independent of circuit | Dependent on circuit |
| | Results in a nonbanded system | Can result in a diagonal system for some types of circuits |
| Irregularity in Circuit | Not an issue | Is an issue |

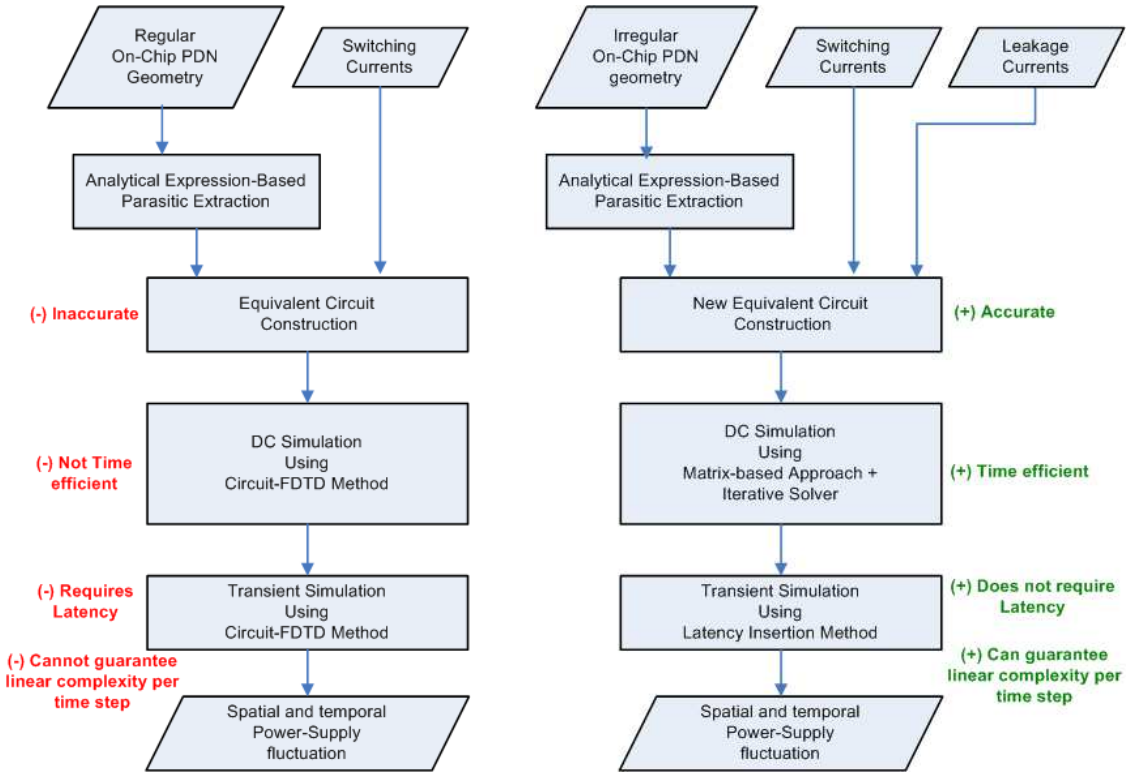
An FDTD-like method for circuits has been traditionally applied to simulate signal propagation in the distributed equivalent circuits of transmission lines. These simulations are useful for signal integrity analysis. Uniform transmission lines are treated in [47] and nonuniform transmission lines are treated in [48]. The latter approach is referred to as the latency insertion method (LIM). LIM is the method proposed in this dissertation for the on-chip power grid simulation. Details regarding LIM are given in Chapter 6.

*Prior Work in Applying an FDTD-like method for Power Grid Problems*

It is to be noted that the power and ground lines in the on-chip power grids are modeled as distributed lossy transmission lines. Taking a cue from the above works [47], [48], an FDTD-like method has been applied to simulate PSN in power grids in [39] and [31]. Both these approaches are collectively termed as a circuit-FDTD method. The simulation for power integrity requires a DC simulation (to setup initial conditions for the transient simulation). This is because the grid is physically connected to a DC power supply. In [39], [40], only the transient simulation is focussed. In [31], [41], the circuit-FDTD method is applied for the DC simulation too. In [39], a frequency-independent transmission line model is used. On the other hand, in [31], a frequency-dependent transmission line model is used. A frequency-dependent model is necessary when the effect of the losses in the silicon substrate are to be included in the PSN simulation. Both these approaches have been applied only for regular power grids. The approach proposed in the prior work [39], [40], [45], [41], [31] is also described in Figure 9(a).

*Limitations of Prior Work based on the FDTD-like Methods - Focus of this Dissertation*

There are some serious limitations in the approaches proposed in [39], [40], [31], [41] (see Figure 9(a)). *The specific objective of this dissertation is to enable an FDTD-like method for on-chip power grid transient simulation without the limitations of [39], [40], [31]*, [41] (see Figure 9(b) for the proposed approach). The following are some




(a) Prior approach [45], [39], [40], [41], [31].

(b) Proposed approach in this dissertation.

**Figure 9. Comparison of the prior and proposed approach in the FDTD-based circuit simulation of PSN in on-chip PDNs.**

of the serious limitations of [39], [40], [31], [41].

1. *These approaches assume that power grid equivalent circuits to be such that a nFDTD-like method can always be enabled.* Specifically, it is assumed that latency is present at every node in the power grid equivalent circuit. It will be shown later in this thesis this assumption is not correct and can lead to inaccurate PSN computation. For example, these approaches assume that every node in the power grid sees an ideal system ground. To this system ground, a capacitance is dropped from the node. It will be shown later in this dissertation (in Chapter 6) that this assumption is not correct. When latency is missing in circuits, the circuit-FDTD method can not be applied. In this dissertation (Chapter 6), an FDTD-like is enabled even in power grid equivalent circuits lacking latency using the LIM.

2. *These approaches do not provide a way to retain the linear computational complexity per time step of the FDTD method in the presence of coupling capacitance.* Coupling capacitors are always present in the power grid equivalent: on-chip decoupling capacitors, and coplanar layer line-to-line coupling capacitors, adjacent layer line-to-line capacitors. When coupling capacitances are present, node voltage update process requires solving a nonbanded system (and not a diagonal system as shown in Figure 8). The size of this system is equal to the number of nodes that are coupled to each other through capacitors. Solving a large nonbanded system using a direct solver is not always a computationally efficient process. Direct solvers are important for accuracy and numerical robustness, which are as important as the complexity. So the linear memory and linear run time requirements of the FDTD-like process may be violated. In this dissertation, a new approach, known as latency insertion method (LIM), has been employed that guarantees linear computational complexity per time step of the transient simulation.

3. *These approaches have not handled DC simulation correctly and efficiently.* In [39], [40], only the transient simulation is focussed. In [41], [31], the circuit-FDTD method is employed for the DC simulation. All these approaches have not addressed the main reason for performing the DC simulation: an average current flowing out of nodes. It will be shown later in this dissertation (in Chapter 4) that it is not time efficient to employ the circuit-FDTD method for the DC simulation. A better thing to do is to employ a SPICE-based solver augmented with an iterative solver instead of a direct solver. Also, in this chapter, transistor leakage current (which has nonzero average value) is included in the PSN simulation; its effect on DC IR drops is shown.

4. *These approaches have not been applied to irregular PDNs.* Power grids are usually regular only in initial stages of the design. During a floorplan stage, they tend to become irregular. In FDTD-based simulations, the numerical stability is also a function of the irregularity of the geometry. Therefore, the performance of these approaches in irregular PDNs has to be demonstrated and proven. In Chapters 2 and 5, the performance of these approaches on irregular PDNs is established.

*Besides, all the prior FDTD method based methods for circuits, whether for signal integrity simulation, including [47], [48], [49], or for power integrity simulations, including [39], [40], [45], [41], [31], [42], have not proven the stability of transient simulations for irregular transmission lines.* This is not an issue for SPICE-based approaches but is an issue for FDTD-based approaches. In this dissertation, the stability of the LIM (a FDTD method-based approach) is proven for inhomogeneous RLC and GLC circuits. The power grid equivalent circuits are based on RLC-type circuits.

In this dissertation, a new equivalent circuit for power grids has been proposed. In this circuit, not all nodes have latency. It has been shown that circuit-FDTD method

can not be applied to these circuits. To enable an FDTD-like scheme, latency insertion method is used. In this method, latency is inserted in a node when it is missing. It has been shown that inserting this latency is crucial in retaining the computational complexity of the FDTD method. Most of the objectives of this research are met through an LIM formulation. The proposed formulation has one limitation: the upper bound for the time step of the transient simulation is a tiny value. This makes the total time taken for the transient simulation large. The estimated time complexity of the overall time complexity is $O\left(N_n^{2-2.5}\right)$. It is expected that this time complexity may be reduced to $O\left(N_n\right)$ through ADI-based methods.

#### *1.2.2.3 Conclusions*

Based on the literature survey, the following conclusions are made for PSN simulation in on-chip PDNs:

1. Most of the existing techniques for on-chip PDN simulation are based on implicit numerical integration rules. When an implicit integration rule is used, the time step of the transient solution does not have an upper bound, and the transient solution is unconditionally stable. Irregularities in on-chip PDN geometries do not pose a problem during the simulation when implicit integration rule is used. In techniques that use implicit integration rules, a large ($\geq$ 1 million nodes), sparse and nonbanded system of equations has to be solved. Solving such large systems accurately and efficiently is a challenging problem computationally. Direct solvers based on the Gauss elimination are the most accurate and do not give rise to convergence problems. However, they are not computationally efficient when applied to nonbanded systems when no attention is paid to the node ordering. The efficiency of these solvers can only be improved with sophisticated node renumbering schemes (See [50]).

2. Most of the prior approaches to power grid simulation replace the direct solver

with either an iterative solver or a statistical solver. Approaches with these new solvers improve the computational complexity, but can degrade the accuracy and/or the robustness of the solution. Direct solver based approaches are preferable for the transient simulation. Hierarchical solvers that use direct solvers compromise accuracy. ADI-based solvers that also use direct solvers guarantee optimal complexity without compromising too much accuracy. However, their performance, specifically their unconditional stability and accuracy, for analyzing irregular PDNs have yet to be understood.

3. Only a small number of the existing techniques are based on finite-difference formulation. These techniques apply an FDTD-like procedure for the power grid circuit simulation. These techniques inherit the advantages of the FDTD method in only a few type of equivalent circuits and inherit the disadvantages of the FDTD method in all types of equivalent circuits. These techniques require solving only a system with diagonal matrices; the solution to such a system can be obtained accurately and efficiently by directly inverting the matrix. Hence, these techniques are free from the problems that plague the techniques based on implicit integration rules. The transient results from an FDTD-like formulation are as accurate as the results from direct solvers. In these techniques, the memory complexity of the transient solution is $O(N_n)$, and the time complexity of the transient solution is $O(N_n)$ per time step. These techniques can handle nonlinear sources efficiently. These techniques have the following drawbacks: 1) They preserve linear complexity per time step of the simulation only in certain types of equivalent circuits. They assume that the equivalent circuits of power grids to be in one of these types. However, this assumption is not correct, and hence the linear computational complexity per time step of the simulation is no longer guaranteed. 2) DC simulation using these techniques is not time efficient. 3) They have not been applied for irregular power grids. 4) Stability

of these techniques for irregular power grids have not been proven. 5) $\Delta t$ of the transient solution has an upper bound that is small (in femtoseconds). Because $\Delta t$ is small, $N_t$ is large. As a result, the time complexity for the transient solution is worse than $O(N_n)$. Methods such as the ADI method to relax the time step constraint have not been applied thoroughly to these techniques.

4. The circuit models for the on-chip PDN used in most of the existing techniques are simple and approximate. On-chip PDN analysis using more accurate circuit models have not been addressed thoroughly.

5. Most of the existing techniques do not include the effect of the lossy silicon substrate.

## 1.3 Need for Causal Transient Simulation of Interconnects Characterized by Band-Limited Frequency Domain Data and Terminated by Arbitrary SPICE circuits

From Section 1.2, it can be observed that all on-chip PDN simulators are based in the time domain. These simulators work with SPICE circuits and produce transient voltages as outputs. On the other hand, from Section 1.1, it can be observed that majority of the package PDN simulators are based in the frequency domain. These simulators work with the full-wave electromagnetic model of the PDN (or sometimes with SPICE circuits) and produce frequency responses (mostly impedance as function of frequency) as outputs. It is necessary to simulate the on-chip PDN together with the package PDN. The common approach in on-chip PDN simulators is to model the package as a SPICE circuit and perform a time-domain simulation on the combined SPICE circuit. Considering the large problem sizes already encountered in the chip, this approach only compounds the problem. Besides, this approach requires knowing the geometry particulars of the package, which might not always be available to chip simulators. Sometimes, only the impedances looking into the package from C4 bumps

are available as a function of frequency. These impedances are usually available up to a certain maximum frequency. These impedances can be obtained either from measurements or from an electromagnetic solver. These impedances are presented as a matrix with size equal to the number of observation points. This number can be same as the number of C4 bumps. Therefore, multiport b.l.f.d. data are available for the package PDN. In these times, it might be necessary to perform a transient simulation of the on-chip PDN characterized by SPICE circuits with the package PDN characterized by frequency-domain data.

Such a transient simulation is fundamentally different from the simulation focussed by existing on-chip PDN simulators. Simple property of a transient simulation like the causality of transient results may not be automatically met in this new transient simulation. On the other hand, this property is not an issue in simulators dealing only with SPICE circuits (with non-negative component values). Causality is critical for an accurate simulation (reasons described later in this section). Therefore, there is a need for developing a time-domain technique for the causal transient simulation of interconnects characterized by b.l.f.d. data with SPICE circuits (see Figure 10). Apart from causality, there are two other desired features in this new simulator: 1)

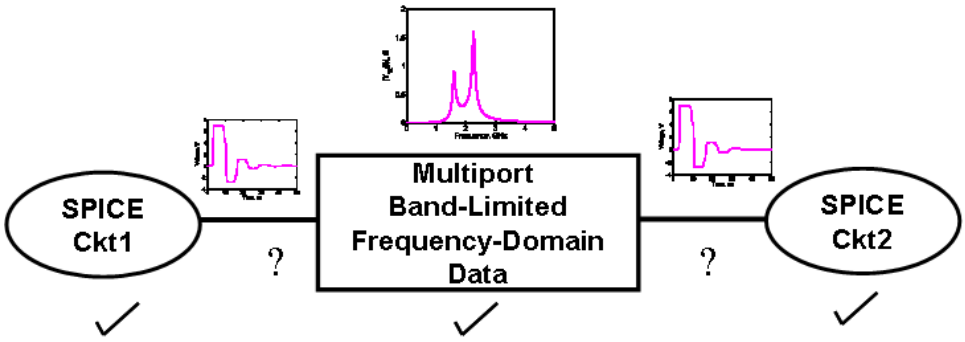



**Figure 10. Transient simulation of band-limited data with SPICE terminations. The quantities above tick marks are given as inputs, while the quantities above the question marks are to be determined.**

There should not be any restrictions on SPICE circuits, as there are many different equivalent circuits for an on-chip PDN. 2) The simulation should preferably handle

a large ($> 100$) number of ports. As a port in this simulation corresponds to a C4 bump, the number of ports is usually in thousands. Assuming that bumps are collapsed together (only for modeling purposes), at least an ability to handle hundreds of ports is desired.

Fortunately, a need for this transient simulation framework has already been felt in signal integrity simulations; the theory behind this framework is being developed since early 1990's. In signal integrity simulations, this framework is necessary to analyze lossy transmission lines whose scattering parameters are known and are terminated by linear/nonlinear drivers/load. This framework is, in general, necessary for simulating signal propagation in any interconnect (not just transmission lines) characterized by frequency-domain data and terminated by SPICE circuits.

Unfortunately, most of the existing approaches in this simulation framework are not computationally efficient when the number of ports is large ($> 100$ ports). Among the approaches that are efficient, causality of the transient results are not always ensured. Among the approaches that are efficient and causal, either the accuracy is compromised or arbitrary terminations can not be handled. Therefore, there is a need to develop a time-domain technique that is accurate and computationally efficient, ensures causality, and handles arbitrary termination for the transient simulation of interconnects characterized by band-limited data and terminated by SPICE circuits.

## 1.4 Causal Transient Simulation with Band-Limited Frequency-Domain Data with SPICE Terminations

Techniques for the transient simulation of interconnects characterized by b.l.f.d. data are fundamentally different from a technique for a traditional circuit simulation: the techniques for the former simulation do not have an ordinary differential equation or a partial differential equation model of the interconnect. Such a model is crucial in traditional circuit simulators, which solve only differential equations. When only

frequency-domain data are known, the electrical quantities (e.g., voltage, current, etc) at the end points, referred to as ports, of the interconnect are related to each other in the time domain through convolution. These techniques therefore require solving the convolution relations among port quantities with additional conditions on these quantities from the terminations at ports.

**Stability and Causality of Transient Results**

As with any transient simulation, it is important to make sure the transient results are stable and causal. Stability of results means that the magnitude of transient results is always (i.e., for all time) bounded when the magnitude of input is bounded. Causality of results means that the output does not precede the input (see [51], [52]). This form of causality is referred to as primitive causality in [51]. This form of causality does not account for the propagation delay experienced by the input before it reaches the (physical) location of the output. When the output is present only after the input reaches the location of the output, then transient results are said to be delay-causal. This new form of causality is referred to as relativistic causality in [51]. Long transmission lines have a large propagation delay that can not be ignored. Therefore, it becomes important to make sure transient results are not only causal but also delay-causal.

**Stability and Causality Depends on Data**

These requirement on transient results are dependent only on the system, i.e., on the frequency-domain data. Stability of transient results are ensured if the frequency-domain data are passive. If the data are scattering parameters, then for the data to be passive, the singular value of the scattering matrix should be less than or equal to unity [53]. Similarly, for the data to be causal, the real and imaginary parts of a port-to-port frequency response cannot be independent of each other and have to be related through an Hilbert transform [51]. In [54], this dependence has been made use of to verify causality of the f.d. data.

**Stability and Causality may be not preserved by Techniques handling data**

Though the frequency-domain data are passive and causal, transient results can still not be stable or causal if the data are known only for a limited bandwidth and not for an infinite bandwidth [55], [56], [57]. The reason has to do with passivity and causality properties of the multiport impulse responses corresponding to the data. To completely determine the inverse fourier transform of a frequency response, the frequency response should be known up to a frequency after which the response is zero. As the frequency response is not known above a certain frequency (because of the limited bandwidth available), all techniques for this transient simulation make assumptions about the behavior of the response from the frequency up to which the frequency response is known to infinite frequency. Some kind of assumption is once again needed if the data are not known for frequencies all the way down to zero frequency. Given that the data are going to be known only in a certain bandwidth and not outside it, the above assumptions are unavoidable. Unfortunately, the multiport impulse response computed (with the assumptions) may not be the actual multiport impulse responses. These modified impulse responses (or their fourier transforms) may not satisfy passivity or causality property. Because these modified impulse responses are employed in the transient simulation, transient results may also not be stable or causal. It is therefore important that the techniques preserve the basic properties of the original data even in the modified impulse responses.

**Other Essential Features Needed of the Technique**

The numerical techniques for the transient simulation should handle arbitrary port terminations and should do so with ease. The arbitrariness of terminations is essential as terminations can be any SPICE circuit. The techniques should also be computationally scalable with respect to the number of ports. The computational scalability w.r.t. number of ports is required in applications that deal with large (>

100) number of ports. One example of such an application is presented in [56], [57]. Finally, these techniques should yield accurate transient results. Maintaining a good accuracy is important, as all the techniques finally work only with modified responses and can naturally comprise a bit of accuracy.

### 1.4.1 Specific Objectives of This Dissertation

The specific objectives of this part of the dissertation is to develop a time-domain technique for the transient simulation of interconnects characterized by b.l.f.d. data with the following features:

1. The technique should be able to handle a large ($> 100$ ports) number of ports.
2. The technique should ensure the causality of transient results.
3. The technique should handle arbitrary port terminations.
4. The technique should ensure reasonable accuracy.

### 1.4.2 History of Prior Work and Its Limitations

The transient simulation with b.l.f.d. data consists of two steps:

1. The multiport f.d. data are converted to time-domain multiport impulse responses. This step controls the accuracy, stability, causality, and efficiency of the overall transient simulation. Needless to say, this step is important.
2. In the second step, port voltages and/or currents are computed from the impulse responses (from the first step) and the port terminations: The multiport impulse responses relate port quantities, such as port voltages and port currents, through convolution. The port terminations enforce an independent set of conditions between the voltage and the current at a port. This step, therefore, involves solving the convolution relations with the termination conditions. The ease with which arbitrary port terminations are handled depends on how the termination

conditions are constructed and solved with the convolution relations. This step also affects the accuracy and efficiency of the simulation.

Depending on how the first step is performed, the existing methods to this transient simulation can be broadly categorized into one of the following two approaches.

#### *1.4.2.1 Recursive Convolution-based Approaches*

In the first approach, referred to as the recursive-convolution-based approach [58]–[59], a differential equation representation of the data is sought. Such a representation makes the second step of the transient simulation simple. A differential equation representation is obtained in the time domain by approximating the f.d. data by matrix rational functions in the frequency domain. Once the rational functions for the data are computed, these functions can be converted to time-domain impulse responses through an inverse laplace transform. These impulse responses are usually exponentials in time. This transform is performed analytically, as the poles and residues of the rational functions are known already. Since the poles are known, the convolution can be performed recursively, which scales $O(N_t)$ in time complexity, where $N_t$ is the number of time steps. The computational complexity of the rational-function fitting depend on how many port-to-port responses are fitted simultaneously and on how many poles are required for an accurate fitting of the data. When $N_p$ port-to-port responses (out of the total $N_p^2$ responses) are fitted together, the memory and time complexities scale as $O(N_p^2 N_{pl}^2)$ and $O(N_p^4 N_{pl}^3)$, respectively, where $N_p$ refers to the number of ports and $N_{pl}$ to the number of poles required to fit $N_p$ port-to-port responses. Therefore, the computational complexity of this fitting can get exorbitant when either $N_p$ is large or $N_{pl}$ is large or both. It is to be noted that the optimal memory and time complexities of the first step are $O(N_p^2 N_f)$ each, where $N_f$ is the number of frequencies for which the data are known. Because of the approximation involved in the fitting procedure, in some cases, the accuracy of the transient simulation can be poor.

Rational function fitting procedure does not guarantee that the fitted functions are passive. Passivity has to be explicitly enforced on them. In [52], the conditions on a frequency response to be passive are described. If rational function system is passive, then the system is causal too [51], [52]. A causal impulse response makes the transient results causal too. However, in these approaches, the propagation delay is not automatically captured [60], [57]. The propagation delay has to be ensured explicitly [59], [61].

#### *1.4.2.2 Numerical Convolution-based Approaches*

In the second approach, referred to as the numerical-convolution-based approach [62]–[56], no assumptions are made about the form (i.e., whether exponentials or not) of the impulse responses like in the recursive convolution. Instead, the first step is accomplished numerically through a simple inverse fast fourier transform (IFFT) of the f.d. data. Owing to the IFFT, a numerical convolution is employed for the transient simulation. The time complexity of this convolution scales as $O(N_t^2)$, where $N_t$ is the number of time steps in the transient simulation. This complexity can be alleviated to $O(N_t \ln N_t)$ through fast convolution methods [56]. The memory and time complexities of the first step scale as $O(N_p^2 N_t)$ and $O(N_p^2 N_t \ln N_t)$, respectively, where $N_t$ refers to the number of time steps. These complexities are close to optimal values for this step, which is $O(N_p^2 N_t)$ each for memory and time. It is to be noted that these complexities are independent of the nature of the f.d. data and, therefore, are also independent of $N_{pl}$, unlike the recursive-convolution-based approach. *In this dissertation, owing to the computational effectiveness of the IFFT procedure, numerical-convolution-based approach has been adopted.*

*Noncausality because of Band-Limited Data*

When impulse response is obtained through IFFT from a band-limited data, the impulse response is not always causal. This noncausality is because of the band-limited nature of the data: the computed impulse response is the convolution of

the actual impulse response (which can be causal) with a sinc function (which is noncausal) and can therefore be noncausal. Note that the convolution with the sinc function is the artifact of band-limiting (≡ multiplying the frequency response by a gate function of finite bandwidth). The band-limited nature of data also makes it difficult to capture the propagation delay in the impulse responses.

*Ensuring Causality is Important For Ensuring Accuracy !*

A noncausal impulse response is not desirable in any convolution-based (recursive or numerical) transient simulation. All transient simulation schemes are designed with an implicit assumption that the impulse responses are causal. For example, all convolution-based schemes work with the following definition of convolution:

$$y(t) = \int_{\tau=0}^{t} h(\tau) x(t-\tau)\, d\tau, \tag{1}$$

where $h(t)$ is the impulse response, and $x(t)$ and $y(t)$ are the input and outputs (e.g., port voltages) of an interconnect. The input $x(t)$ is causal. From (1), it can be observed that all the schemes work only with the causal part of the impulse response, i.e., with $h(t)$ for $t > 0$. If the $h(t)$ is noncausal, then the transient simulation, unfortunately, only uses the causal part of $h(t)$. The transient results obtained with a noncausal $h(t)$ is not accurate. Moreover, energy of the output with a noncausal $h(t)$ may not agree with that of the input, which is definitely undesirable. These two points are addressed later in this dissertation (see Chapter 11).

Among the prior numerical-convolution-based approaches [62], [63], [64], [65], [55], [56], the references [62]– [65] do not capture the propagation delay when only the f.d. data are known about the interconnects. These approaches can not also guarantee the causality of the impulse responses. The approaches [55], [57] address specifically the delay-causality problem of the transfer responses (responses between two different ports) (see Figure 11(a)). These approaches extract the propagation delay from the data, compute the impulse response (which can be noncausal) from the data using

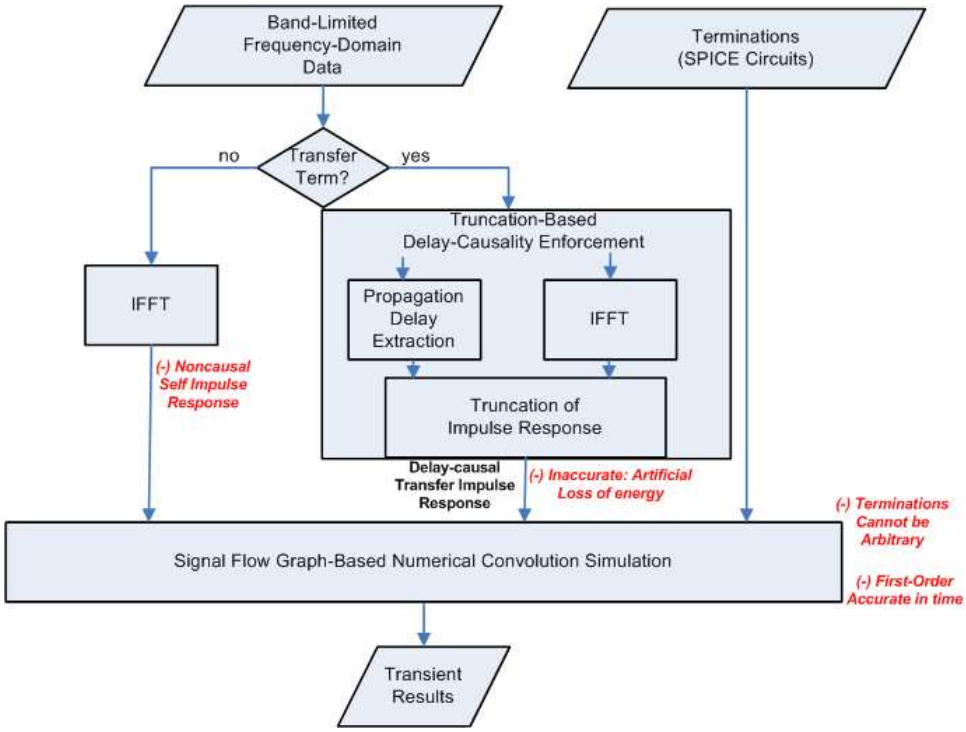
the IFFT, and enforce this delay in the transfer impulse responses by truncating (i.e., zeroing) the part of the impulse response for times less than the propagation delay. Because of this truncation, the transfer impulse response are made delay-causal. Using these delay-causal impulse responses, transient simulation is performed using a signal flow-graph-based approach. This approach essentially addresses the second step of the transient simulation with band-limited data. The approaches [55], [56], [57] apply their delay-causality solution to signal integrity simulation problems.

*Limitations of Prior Work - Focus of this Dissertation*

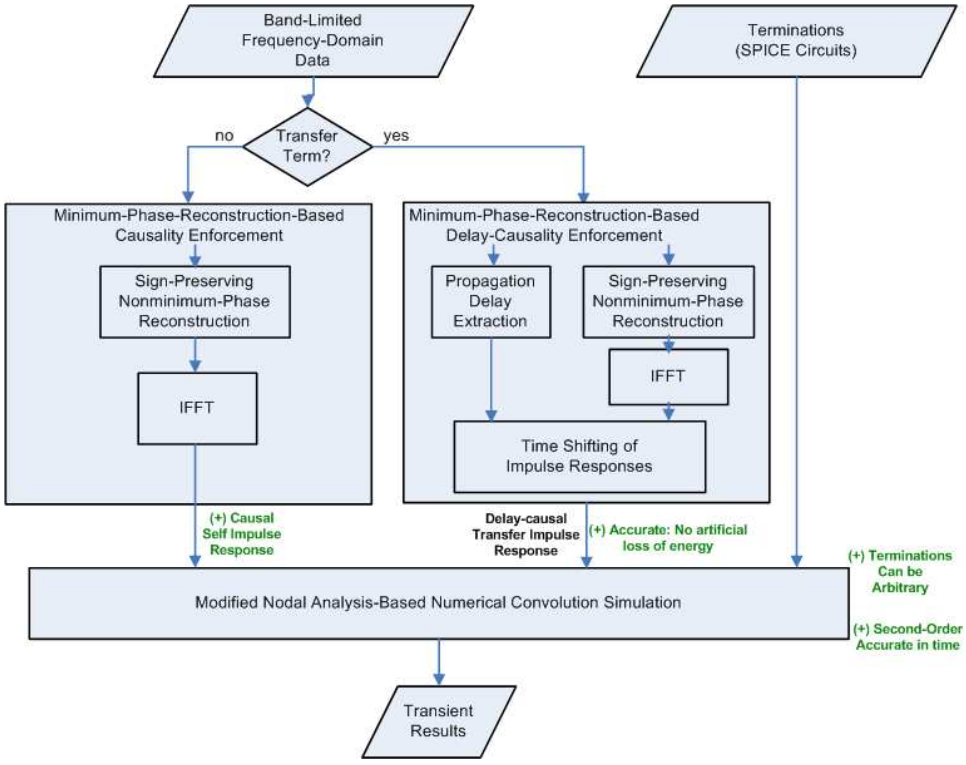
The approaches [55], [56], [57] have serious limitations that can affect the accuracy and capability of the transient simulation (see Figure 11(a)). The objective of this dissertation is to perform the transient simulation of band-limited data for signal integrity applications without the limitations in [55], [56], [57] (see Figure 11(b)). Some of the serious limitations of [55], [56], [57] are as follows:

1. *The fact that the approaches [55], [56], [57] truncate the nondelay-causal part of the impulse response introduces the same kind of inaccuracy described earlier in this section (see description regarding (1)).* Such a truncation can lead to significantly inaccurate transient result. In Chapter 9, the effect of delay causality enforcement strategies on the accuracy is evaluated. Also in this chapter, a new delay-causality enforcement procedure is proposed without the inaccuracy issues of [55], [56], [57].

2. *The transient simulation framework used by approaches [55], [56], [57] are based on signal flow graphs. With a signal flow graph-based (SFG-based) simulation, it is difficult to handle arbitrary port terminations with ease.* The reason is because the SFG-based approaches work only with the input impedance of the terminations even though SPICE circuits are available for the terminations. Computing input impedance of an arbitrary circuit can get difficult. In Chapter 10, this limitation of the SFG-based approaches is discussed more in detail. Also




(a) Prior approach [55], [56], [57].




(b) Proposed approach in this dissertation.

**Figure 11. Comparison of the prior and proposed approach in the numerical-convolution-based causal transient simulation of band-limited frequency-domain data.**

in this chapter, a new transient simulation algorithm based on modified nodal analysis framework has been proposed. With this approach, only the circuit realization of the terminations is dealt with.

3. *The approaches [55], [56], [57] address only the causality and delay-causality of the transfer frequency responses and not those of the self frequency responses.* The band-limitedness of data induces noncausality in both the transfer and self responses alike. Therefore, the approaches [55], [56], [57] work only with noncausal self responses. The undesirable effects of working with a noncausal impulse response is already described. In Chapter 11, this limitation is described more in detail. Also in this chapter, causality of the self responses are ensured through a variant of the causality enforcement technique proposed for transfer responses in Chapter 9.

4. *The approaches [55], [56], [57] can only guarantee a transient simulation whose accuracy is $O(\Delta t)$, where $\Delta t$ is the time step of the simulation.* The reason has to do with the second step of the transient simulation: how the convolution part is integrated with the termination part during the transient simulation. The approaches [55], [56], [57] do not provide a way to improve the accuracy beyond $O(\Delta t)$. It is to be noted that the traditional circuit simulators can guarantee better accuracy. For example, using a trapezoidal integration rule, $O\left((\Delta t)^2\right)$ accuracy can be accomplished. In this dissertation, the transient simulation proposed in Chapter 10 can guarantee accuracy comparable to that of a traditional circuit simulator.

5. *Finally, the approaches [55], [56], [57] do not address the effects of frequency-domain windowing on causality.* It has been shown in this dissertation, *for the first time*, that the frequency-domain windowing is one of the causes for causality violations in numerical convolution-based approaches. This effect of

windowing is described in Chapters 9 and 11. Also in this chapter, the significant inaccuracy a truncation-based causality enforcement can cause in the presence of windowing has been demonstrated. With the new causality enforcement strategy proposed in Chapter 9 (as part of this dissertation), even the causality violations induced by frequency-domain windowing can be dealt with without the inaccuracy issues associated with the truncation-based enforcement strategy.

Finally, some of the difficulties associated with the proposed technique are also addressed. For example, the causality enforcement strategy proposed in Chapter 9 does not model the leading negative sign of the frequency responses. Not modeling a sign of a frequency response can affect the accuracy of the transient results significantly. In Chapter 10, this issue is dealt with in detail. Also in this chapter, a sign-preserving causality enforcement strategy is proposed.

#### *1.4.2.3 Conclusions*

The following conclusions can be made for the transient simulation of b.l.f.d. data:

1. Most of the approaches use a recursive convolution formulation. This requires constructing reduced-order rational functions that fit the port-to-port frequency responses. This fitting process can be computationally inefficient for a large number of ports and/or for a large number of poles.

2. The other approaches use a numerical convolution formulation. Numerical convolution-based approaches require computing the IFFT of the frequency-domain data. The advantages of this approach are that it is accurate, and it can handle a large number of ports. However, with band-limited data, the causality of the simulation can be violated.

3. Most of the numerical-convolution-based approaches do not ensure the causality of the impulse responses. The existing techniques that address causality of

the impulse responses compromise the accuracy of the simulation significantly. Also, these techniques address the causality of only the transfer responses and not the self responses. Noncausal responses affect the accuracy of the transient simulation. Also, the techniques that address causality can not handle arbitrary port terminations with ease. Finally, the effects of frequency-domain windowing on causality have not been clearly understood.

## 1.5 Proposed Research

The objectives of this Ph.D. research are twofold: 1) To develop accurate, numerically robust, and computationally efficient time-domain algorithms to compute PSN in on-chip PDNs with irregular geometries. 2) To develop accurate and computationally efficient time-domain algorithms for the causal cosimulation of band-limited frequency-domain data with SPICE circuits.

The time-domain PSN simulation involves two steps: 1) a DC simulation to set the initial conditions and 2) a transient simulation to launch and find the effects of the switching sources. For the transient simulation, the latency insertion method (LIM) has been proposed (see Figure 9(b)). The advantages of the proposed method is that its accuracy scales as $O((\Delta t)^2)$, 2) its memory complexity is $O(N_n)$, 3) its time complexity is $O(N_n)$ per time step, and 4) nonlinear sources can be simulated efficiently, where $\Delta t$ is the time step, and $N_n$ is the number of nodes in the equivalent circuit. The disadvantages of the proposed method are that the time step, $\Delta t$, has an upper bound and is usually small. The DC simulation using the LIM has been found to be time inefficient. Therefore, the DC simulation has been performed by solving a sparse matrix using an iterative solver (see Figure 9). The new technique has a better run time compared to the LIM and has an $O(N_n)$ memory complexity. The PSN has been computed in the presence of irregularities in power-ground lines of the on-chip PDN. The proposed simulation takes into account the switching and leakage

currents, the crossover and on-chip decoupling capacitances, and package parasitics.

For cosimulating frequency-domain data with SPICE circuits, a new numerical convolution-based approach has been proposed (see Figure 11(b)). The proposed approach has the following advantages over the prior approaches: 1) The proposed approach is scalable for a large number of ports, unlike recursive convolution-based approaches. 2) The proposed approach can produce more accurate transient results than recursive-convolution-based approaches in some transmission-line simulations. 3) The proposed approach ensures the complete causality of the transient results by enforcing causality of both the self and transfer responses. The prior numerical convolution-based approaches either do not enforce causality or do enforce it only incompletely. 4) The proposed approach ensures causality using a new causality enforcement procedure, based on a minimum-phase/all-pass decomposition of the frequency responses. This procedure has been shown to be more accurate than the traditional truncation-based causality enforcement procedures (See Figure 11).

## 1.6 Completed Research

The following work has been completed:

1. *Identification of Accuracy and Efficiency Issues in Performing a DC Simulation using Circuit-FDTD Method (Chapter 2)*

   A new problem has been identified when the circuit-FDTD method is applied to DC simulation. This problem concerns the accuracy of the DC node voltages computed. It has been found that when DC simulation is computed using the circuit-FDTD method, the oscillations from step responses do not settle down (or die down) in some nodes. When transient PSN simulation is started on a circuit with unsettled step responses, the PSN can be computed inaccurately in two ways: 1) The PSN can have contributions not only from switching currents (which it should) but also from unsettled step responses (which it should not).

2) The PSN can be observed in a location even before the effect of the switching current can be felt at the location, i.e., the PSN computation can violate causality. This new problem has been solved by running the DC simulation for sufficiently long time so that the step responses are significantly settled in all nodes. Unfortunately, in the modified simulation, it has been observed that the DC simulation took majority of the total simulation time.

2. *Time-Efficient DC Simulation (Chapter 4)*

   To improve the run time of the DC simulation, DC node voltages are not computed by the circuit-FDTD method; instead, they are computed by solving a sparse linear system arising out of the modified nodal analysis (MNA) of the circuit using an iterative method. It has been found that the new method has linear memory complexity and has a better run time compared to the circuit-FDTD method.

3. *Efficient Reformulation of Circuit-FDTD Method with Crossover Capacitance (Chapter 3)*

   The overlap capacitance between power-ground lines in adjacent metal layers, also known as the crossover capacitance, has been included in the on-chip PDN equivalent circuit. The formulation of the circuit-FDTD method has been modified to include the crossover capacitance. This new formulation requires solving a small matrix system at each time step, for finding the voltages of the nodes that are capacitively coupled. The size of this matrix system has been shown to be upper bounded by the number of metal layers in the chip. Therefore, the memory and time complexities of the overall simulation scale as $O(N_n)$ and $O(N_t N_n)$, respectively. This new formulation has also been extended for the frequency-dependent on-chip PDN equivalent circuit. Prior circuit-FDTD approaches either have not modeled this capacitance or have not shown that the

linear computational complexity per time step of the transient simulation can be guaranteed.

4. *Circuit-FDTD Method for Irregular On-Chip PDNs (Chapters 2 and 5)*

   The circuit-FDTD method, originally applied to regular (uniform line spacing, uniform line widths, and continuous power/ground lines running from one side of the chip to the other) on-chip PDNs, has been extended to on-chip PDNs with nonuniform line spacing. The accuracy of the implementation has been verified through simulations.

   The PSN has also been computed in on-chip PDNs with nonuniform line widths and with broken lines. It has been found that in on-chip PDNs with broken (or discontinuous) lines, finding the locations of vias and crossover capacitors is computationally more challenging than it is in on-chip PDNs with unbroken (or continuous) lines. Finding via (and crossover capacitance) locations required computing projections of all lines (on the others) in adjacent metal layers. The time complexity of finding line projections scales as $O(N_l^2)$, where $N_l$ is the number of power-ground lines in the on-chip PDN. As for the circuit-FDTD method, because of irregularities in the PDN, each node in the PDN places a separate constraint on how large a time step should be for stable results. The time step of the transient simulation should be chosen as the smallest among the different time steps.

5. *On-Chip PDN Simulation using LIM (Chapter 6)*

   Circuit-FDTD method guarantees linear computational complexity per time step of the transient simulation only in circuits where there is latency in every node and branch of the circuit. It is shown, however, that this latency requirement may not be met in equivalent circuits of on-chip PDNs. To preserve the computational complexity, it has been proposed to insert artificial

latency in missing places of the circuit. The circuit-FDTD method augmented with artificial latency is referred to as the latency insertion method (LIM). LIM, like any FDTD-based method, is only conditionally stable. The time step of the transient simulation can not be arbitrary and depends on the smallest inductance-capacitance in the circuit. Care has to be taken about the values of artificial latency elements. If the artificial element values are too large, then the accuracy can be significantly affected. On the other hand, if these element values are too small, then time step of the transient simulation has to be made small. Unlike LIM, which rely on several (transient simulation) iterations to compute the latency elements, this dissertation proposes closed-form expressions for computing the latency elements. These expressions take into account the element values, the maximum frequency in the excitation, and the accuracy required. Therefore, time step can be made just small enough to meet the accuracy requirements. Unlike in LIM, simulation need not be repeated for accuracy. The LIM-enabled power grid transient simulator is demonstrated to be as accurate and robust as SPICE, to have linear time complexity per time step of the transient simulation, and to have linear memory complexity for the whole transient simulation. The total number of time steps required in this simulator is shown to be $O(N_n^{1-1.5})$ for practical on-chip power grid problems.

6. *On-Chip LIM Including On-Chip Decoupling Capacitors and Package Parasitics (Chapter 7)*

   LIM has been extended to simulate PSN in on-chip power grids in the presence of on-chip decoupling capacitors. An RC model has been used for the on-chip decoupling capacitance. To retain optimal memory and time complexity per time step of the simulation, to each on-chip decap, a fictitious inductance has been inserted. The accuracy of the simulation has been verified against SPICE. The effect of the on-chip decoupling capacitance on the PSN has been

demonstrated.

LIM has also been extended to simulate PSN in on-chip power grids in the presence of package parasitics. The package has thus far been modeled as an ideal voltage source. As a first-level model, the C4 bump and the package has been modeled as a series RL branch. This branch is put in C4 locations. The value of the resistance and inductance can be obtained from the input impedance seen from C4 terminals to the end of the package. The computational complexity of the LIM has been retained. The accuracy of this simulation is verified against SPICE. The importance of modeling package PDNs even while simulating PSN in power grids is verified through simulations.

7. *Effect of On-Chip Inductance on PSN (Chapter 7)*

   A tiny minority of the prior work has studied the effect of the on-chip grid inductance on PSN. Among these, some of them make a case for including the inductance and some of them make a case for the opposite. These prior approaches have been only using PEEC-based modeling, which is inefficient in a pre-layout-level simulation.

   Using the proposed formulation and equivalent circuit, the effect of the on-chip inductance on the PSN has been studied. It has been found that the on-chip inductance has three effects that can potentially affect the PSN computation: 1) On-chip inductance lowers the frequency of the chip-package resonance. 2) On-chip inductance lowers the magnitude of the peak impedance, usually observed near the chip-package resonant frequency. 3) On-chip inductance introduces new resonances at frequencies greater than the chip-package resonant frequency. These extra resonances introduce a fast variation to the PSN. This variation can make the power supply fluctuate beyond the allowed margin, although this violation is only temporary. This sudden variation in power supply would not

be captured if the on-chip inductance is not modelled as part of the power grid simulation.

8. *Stability of LIM for Irregular Power Grids (Chapter 8)*

   LIM, unlike the SPICE-based approaches, is not guaranteed to be stable when there are discontinuities in circuits. Until now, it has not been possible to prove the stability of LIM for inhomogeneous circuits. In this dissertation, the stability of LIM has been proven for inhomogeneous RLC and GLC circuits. With this proof, the proof for stability of LIM-enabled power grid simulation for irregular power grids is established for cases where capacitive coupling can be ignored. Moreover, analytical stability conditions in the form of a Courant-like time step have been derived for inhomogeneous RLC and GLC circuits.

9. *Conditional Stability of Alternate Direction Implicit Methods*

   Alternate direction implicit (ADI) method has been used to relax the time step of the transient simulation using the transmission line method (TLM), an explicit method similar to the circuit-FDTD method. It was found that 1) the ADI method can only be applied to mesh-type equivalent circuits (where two orthogonal directions of propagation are possible in every metal layer) and 2) the ADI method for mesh-type equivalent circuits becomes unstable for some choices of time step when open-circuit boundary conditions are applied at the circuit boundary. Therefore, it has been concluded that the ADI method can not be used to relax the time step of the circuit-FDTD for the on-chip PDN equivalent circuits considered in this research.

10. *Delay Causality Enforcement Using Minimum-Phase/All-Pass Decomposition (Chapter 9)*

    When multiport b.l.f.d. data are present, the delay causality of transfer (i.e.,

between two different ports) impulse responses are ensured traditionally as follows: each transfer impulse response is computed numerically using IFFT, an average propagation delay is extracted next from the frequency response, and the impulse response (after IFFT) is truncated for times less than the corresponding propagation delay. It has been shown that such truncation-based causality enforcement techniques do not preserve the energy of the individual frequency response and can result in inaccurate transient results. To avoid this drawback, a new causality enforcement technique has been proposed. In this technique, each transfer frequency response is causally reconstructed in the frequency domain using a minimum-phase/all-pass decomposition of the frequency response, the reconstructed frequency response is converted to the corresponding impulse response numerically using the IFFT, and this impulse response is shifted in time to account for the propagation delay. It has been observed that the new technique does not suffer from the inaccuracy issues observed in the truncation-based techniques. The accuracy of the proposed method has been verified.

11. *Delay-Causal Transient Simulation of Band-Limited Frequency-Data with SPICE Circuits (Chapter 10)*

    Prior delay-causal numerical convolution-based approaches are based on a signal flow graph-based framework and therefore can not handle arbitrary port terminations. A new delay-causal transient simulation engine that integrates band-limited frequency domain data characterizing a multiport linear system with SPICE circuits has been implemented. This integration has been achieved by formulating a numerical convolution-based approach in an MNA framework. The advantage of this engine are that the port terminations can be arbitrary and the transient results are delay-causal. The accuracy of the transient simulation has been verified for frequency-domain data characterizing transmission

lines.

12. *Sign-Preserving Minimum-Phase/All-Pass Decomposition (Chapter 10)*

    Using the delay-causality enforcement technique discussed thus far, leading signs of frequency responses are not preserved consistently. Not preserving this sign can make the transient results inaccurate. This nonpreservtion of sign is because of the limitation of the existing functional form of the all-pass component. Using this form, a leading negative sign of a frequency response can not be modeled. To capture the leading sign, a constant sign term has been included as part of the all-pass component. The accuracy of the new functional form of the all-pass component and the transient results using this decomposition have been demonstrated.

13. *Causality Enforcement for Self Frequency Responses (Chapter 11)*

    Thus far, all causality efforts have only focussed on transfer impulse responses; each self impulse response is computed as the IFFT of the corresponding band-limited frequency response and is assumed to be causal. This assumption is first demonstrated to be not correct. The traditional methods implicitly truncate the noncausal part. As a result, they introduce the same kind of inaccuracy observed for transfer responses. To overcome this inaccuracy, even the self responses are reconstructed using the sign-preserving minimum-phase/all-pass decomposition. Subsequent IFFT of a reconstructed self frequency response yields a causal self impulse response. Transient simulation with the resulting impulse response does not inaccuracy issues related to the truncation-based technique. The accuracy of the reconstruction and the transient results are demonstrated.

14. *Frequency-Domain Windowing Induces Causality Violations (Chapters 9 and 11)*

In all numerical-convolution-based approaches, the b.l.f.d. data are usually subjected to a frequency-domain windowing for making the transient results smooth and stable (sometimes). The strength of the window is chosen based on accuracy and stability considerations. In this dissertation, it has been shown, for the first time, that frequency-domain windowing makes causal frequency-domain data noncausal. It has been also demonstrated that the bigger the strength of windowing, the larger the noncausality the data become. As a result, when frequency-domain windowing is applied, the transient results are not going to be causal unless ensured.

Band-limited nature of data can be considered as applying a rectangular window to the data. However, this windowing does not make the data causal, instead it only makes the time response noncausal. This noncausality is not so serious as the noncausality from other windowing.

## 1.7 Dissertation Outline

The rest of dissertation is organized into two parts (see Figure 12): In first part, PSN simulation in on-chip power grids using two FDTD-like methods (circuit-FDTD method and LIM) has been described. In the second part, causal transient simulation of band-limited frequency-domain data is described.

The first part consists of seven chapters: Chapter 2 through Chapter 8. In Chapter 2, the performance of the circuit-FDTD method is investigated. In Chapter 3, the previous circuit-FDTD formulation has been reformulated to model crossover capacitances without comprising the linear complexity per time step of the previous formulation. In Chapter 4, an efficient method to perform DC simulation has been described. In Chapter 5, the new transient and DC simulation methods described in Chapters 3 and 4, respectively, have been employed for simulating PSN in power

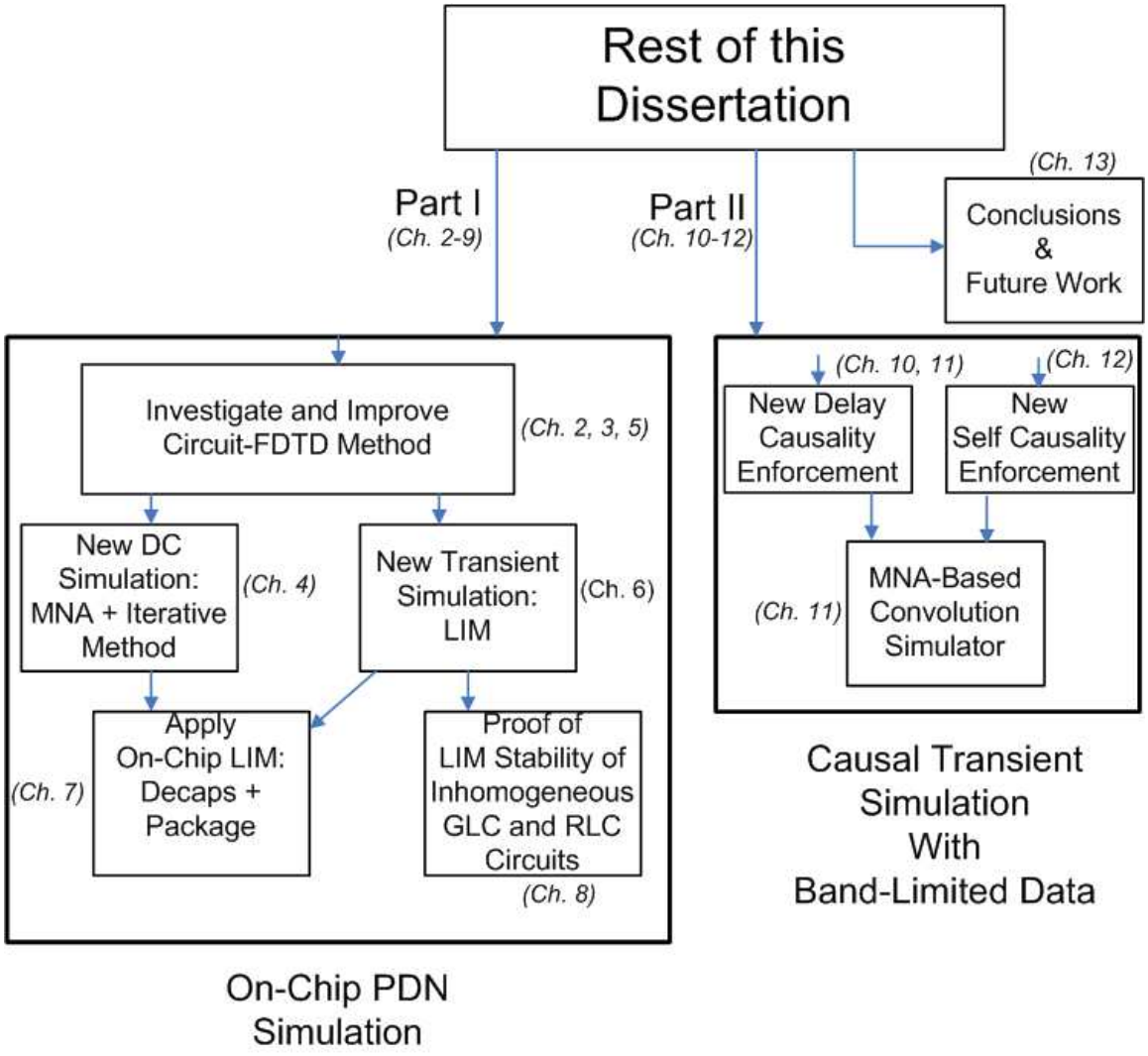



**Figure 12. Organization of the rest of this dissertation.**

grids with various irregularities. In Chapter 6, the transient simulation is reformulated using LIM. The need for this reformulation is described in this chapter. The performance of the LIM-enabled on-chip PSN simulation is demonstrated. In Chapter 7, the transient simulation using LIM has been extended to include on-chip decoupling capacitors and the effect of the package PDN. Also in this chapter, the effect of the on-chip inductance on the PSN has been studied. In Chapter 8, analytical stability condition of the LIM for inhomogeneous GLC and RLC circuits are presented. The stability conditions in turn gives the Courant time step.

The second part consists of three chapters: Chapter 9 through Chapter 11. In Chapter 9, a new method to enforce transfer causality in the transient simulation is proposed. The proposed method relies on reconstructing the transfer frequency responses using minimum-phase/all-pass decomposition. In Chapter 10, a new numerical convolution-based transient simulation technique is proposed for simulating multiport band-limited data with arbitrary port terminations. Also in this chapter, the functional form for the all-pass component described in Chapter 9 is modified to include the leading sign of transfer frequency responses. The need for this modification is justified in this chapter. In Chapter 11, unlike the previous chapters, the need for enforcing causality even for self responses has been justified. The causality enforcement technique described in Chapter 10 has been employed to enforce causality in self responses.

Finally, in Chapter 12, the conclusions, the contributions, and the future work of this dissertation are summarized.

# CHAPTER 2

# INVESTIGATION OF ON-CHIP POWER GRID SIMULATION USING CIRCUIT-FDTD METHOD

## 2.1 Introduction

Until now, the circuit-FDTD method was applied to simulate PSN in only regular on-chip PDNs. However, the geometry of an on-chip PDN remains irregular only at the early stages of its design. Hence, the circuit-FDTD method has to be applied to irregular on-chip PDNs as well. Also, the accuracy vs. complexity trade-off of using a frequency-dependent model over a frequency-independent model has not been clearly understood. Finally and more importantly, the performance of the circuit-FDTD method for a DC simulation has not yet been studied. In this chapter, the concerns mentioned above are addressed. For the irregular power grid simulation, the circuit-FDTD method in [41], [31] is extended to on-chip PDNs with nonuniform line spacing (see Figure 6), which is one kind of irregularity observed in practical on-chip PDNs. This method is extended to handle two other kinds of irregularities in Chapter 5. The contributions of this chapter and organization of rest of this chapter are described at the end of Section 2.2.

## 2.2 Prior Work

### Circuit-FDTD Method For Irregular PDNs

Circuit-FDTD methods for power grid simulation have been proposed in [39], [40], [45], [41], [31]. All these approaches have only focussed on regular power grids. However, power grids are usually irregular. Irregularities in geometry do not change simulation parameters in SPICE-based approaches (see Table 1). They do, however, in FDTD-based approaches. The circuit-FDTD method is only conditionally stable. The stability condition (which is a simulation parameter) depend on both circuit topology and circuit element values. Irregularities in PDN affect both these

parameters. However, the exact dependence is not known. This dependence has been predicted (without proof) in [48]. It becomes important therefore to establish the working of the circuit-FDTD method for irregular power grids.

**Accuracy vs. Complexity Trade-off**

Among these approaches, the approaches [39], [40], [45] do not model the loss in the silicon substrate (see Figure 5 in Chapter 1 for the location of this substrate). Accordingly, these approaches only use a frequency-independent equivalent circuit. On the other hand, the approaches [41], [31] model the loss of the silicon substrate. They do so by employing a frequency-dependent equivalent circuit for power-ground lines in the on-chip PDN. The circuit-FDTD formulation in [39], [40], [45] cannot, however, be used on the equivalent circuits proposed in [41], [31]. Therefore, a new circuit-FDTD formulation has been employed in [41], [31]. This formulation, like the previous formulation, requires only linear computational resources per each time step of the transient simulation. However, the new formulation is a little more complicated and hence would require more computational resources than the previous formulation. It would be useful to assess the gains in the accuracy along with the extra cost incurred in using a frequency-dependent model over a frequency-independent model. This study would be useful to see if a frequency-dependent modeling can be safely avoided. However, this study has not been done yet.

**DC Simulation**

The approaches [39], [40], [45] have only addressed the transient simulation and not the DC simulation. The approaches [41], [31] use the new circuit-FDTD formulation for both the DC and transient simulations. However, circuit-FDTD-based methods are intended to be employed as high-frequency methods, making them suitable only for a transient simulation. Its performance for a DC simulation should be studied. However, this study has not yet been done.

In this chapter, the circuit-FDTD method has been applied to power grids with nonuniform line spacing. The problems in using a circuit-FDTD method for DC simulation have been found. The accuracy vs. cost trade-off in using a frequency-dependent model over a frequency-independent model has been performed. The following are the conclusions:

1. As an example of irregularity, the line spacing is made nonuniform. It has been found that in such an irregular grid, the only change to the circuit-FDTD method is the time step. The time step required for stability is different at different nodes in the grid. The smallest value among them should be chosen as the time step of the transient simulation.

2. When DC simulation is performed using the circuit-FDTD method, there are situations where the PSN can be erroneously computed. This situation cannot be known apriori; therefore, care has to be taken to minimize the impact on PSN. The impact on PSN can be minimized if the simulation is kept on for a long time. Unfortunately, the resulting DC simulation is observed to be not time efficient.

3. Modeling the loss in the silicon substrate is found to be important. However, the computational requirements are at least increased by twofold when a frequency-dependent model is used, instead of a frequency-independent model.

*The contribution of this chapter are*

1. *the application of the circuit-FDTD method for power grids with nonuniform line spacing*[1] *and*

2. *the identification of performance issues of the circuit-FDTD method for a DC simulation.*

[1] **S. N. Lalgudi**, J. Mao, and M. Swaminathan, "Parasitic extraction and simulation of simultaneous switching noise in on-chip power distribution networks," *IEEE conference on Electromagnetic Compatibility,* Mar. 2005, Zurich.

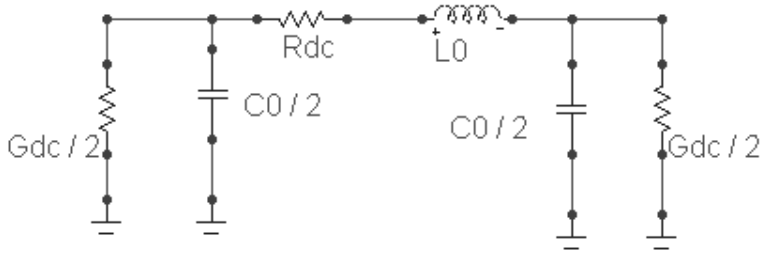



**Figure 13. Frequency-independent $\pi$ - type RLGC model of a single segment of a power/ground line. $R_{dc}$, $L_0$, $G_{dc}$, and $C_0$ are the resistance, the inductance, the conductance, and the capacitance, respectively, at low frequencies.**

The rest of this chapter is organized as follows: In Section 2.3, the equivalent circuit for the PSN simulation in on-chip PDNs is described. In Section 2.4, the formulation of the transient simulation using the circuit-FDTD method is described. In Section 2.5, the DC simulation using the circuit-FDTD method is described. Also, in this section, potential issues in using the circuit-FDTD method for the DC simulation are discussed. In Section 2.6, numerical results are presented. Finally, in Section 2.7, the conclusions of this chapter are summarized.

## 2.3 Equivalent Circuit of Passive and Active Circuits

A simplified 3-D view of a regular on-chip PDN is shown in Figure 5 (see Chapter 1). The power and ground lines in this PDN are modeled as uniform lossy transmission lines. The conducting plane beneath the substrate acts as the reference ground for the transmission line. Each line is segmented further to create additional nodes at the end points of vias. A distributed RLC model can be constructed for power-ground lines. To account for the dielectric loss in the silicon substrate, a conductance term is added to the distributed RLC model. The p.u.l. quantities (R, L, G, and C) of lines are obtained through closed-form expressions described in [31]. In Figure 13, the $\pi$-type equivalent circuit of one segment of the line is shown. Each via is modeled by a lumped series RL circuit. The via resistance (inductance) is obtained as the

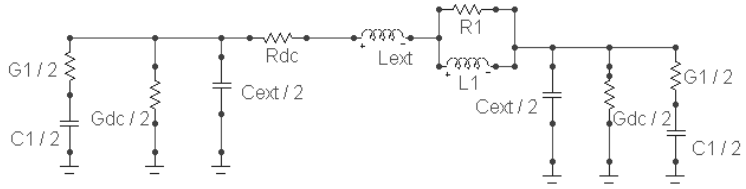



**Figure 14. Frequency-dependent equivalent circuit of a segment of a power or ground line. A first-order Debye model is added to capture the frequency dependency of R, L, G, and C. $R_{dc}$ and $G_{dc}$ are the DC resistance and the DC conductance, respectively; $L_{ext}$ and $C_{ext}$ are the high-frequency inductance (or external inductance) and the high-frequency capacitance, respectively.**

product of the length of the via and the p.u.l. resistance (inductance) of the line in the layer below to which the via is connected. The model shown in Figure 13 is valid when the p.u.l. quantities of the line do not vary with frequency. However, if the p.u.l. quantities vary with frequency, a frequency-dependent model is important. To capture the frequency-dependent p.u.l. parameter variation, each segment of the line in the circuit can be augmented with a Debye model [67]. The new segment in the power/ground line with one Debye term is shown in Figure 14. The parameters ($R_{dc}$, $L_{ext}$, $G_{dc}$, $C_{ext}$, $R_1$, $L_1$, $G_1$, and $C_1$) of the segment can be obtained by fitting the frequency response of the segment to the measured or extracted frequency data using the minimum least squares technique or the vector fitting [68] technique.

When power-ground lines have nonuniform spacing between them, the following changes occur: 1) The loop impedance of power-ground lines is altered and 2) the segment size becomes nonuniform. The procedure for extracting the loop impedance of power-ground lines with nonuniform line spacing has been described in [31]. This procedure is used for computing the PSN in PDNs with nonuniform line spacing. Nonuniform segment sizes cause the time step of the circuit-FDTD method to be different for different nodes.

Until now, the equivalent circuits of only the passive components in the PSN computation have been described. However, the switching sources and the power/ground

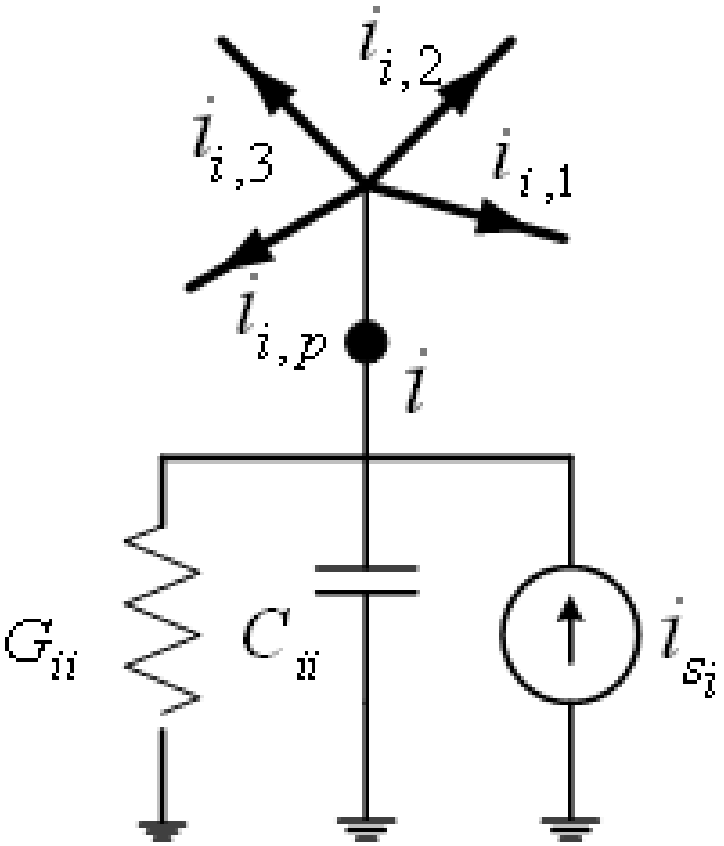
supplies are active components. The switching circuit actually has nonlinear current-voltage characteristics. However, considering the problem sizes encountered in on-chip PDN analysis, the switching source is modeled as a linear source. In this work, the switching source has been modeled as a linear triangular current pulse stream, emulating the periodic switching of a CMOS invertor. The power supply resides on the PCB. However, to manage the complexity of the problem, the power supply is assumed to be present at C4 locations. The power supply is modeled as a DC voltage source.

## 2.4 Formulation of the Transient Simulation

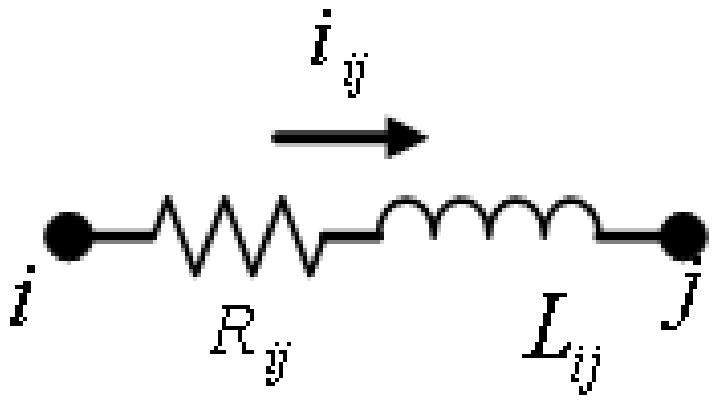
To compute the PSN in the equivalent circuit of the PDN using an explicit method (i.e., FDTD-like method), update expressions have to be derived for node voltages and branch currents. The update expressions for node voltages and branch currents depend on the type of equivalent circuit used for a segment of the power-ground line. Because both frequency-independent and frequency-dependent equivalent circuits can be used for the power-ground lines, the voltage and current update expressions are derived for both types of equivalent circuits.

### 2.4.1 Frequency-Independent Equivalent Circuit

The equivalent circuits observed at a node and in a branch (or segment) in the frequency-independent equivalent circuit of the on-chip PDN are shown in Figure 15. In Figure 15(a), $C_{ii}$ and $G_{ii}$ are the capacitance and the conductance, respectively, between node $i$ and (ideal) ground; $i_{s_i}$ is the current entering node $i$ as a result of the switching circuits; and $i_{i,p}$ is the current entering node $i$ from the $p$th branch connected to node $i$. In Figure 15(b), $L_{ij}$ and $R_{ij}$ are the inductance and the resistance, respectively, of the branch between nodes $i$ and $j$; $v_i$ and $v_j$ are the voltages at nodes $i$ and $j$, respectively; and $i_{ij}$ is the current through the branch between nodes $i$ and




(a) A node




(b) A branch

**Figure 15. A node and a branch in the frequency-independent equivalent circuit of the on-chip PDN.**

$j$. The KCL at node $i$ in Figure 15(a) can be written as

$$C_{ii}\frac{dv_i(t)}{dt} + G_{ii}v_i(t) = i_{s_i}(t) + \sum_{p\,=\,1}^{P_i} i_{i,p}(t). \tag{2}$$

Let $i_{ij}(t)$ and $v_i(t)$ be defined at time instances $n\Delta t$ and $(n+\frac{1}{2})\Delta t$, respectively, where $n = 0, 1, 2, ..., N_t$, and $N_t$ is the total number of time steps. Making use of central difference approximation [69] to represent $\frac{dv_i}{dt}$, a discrete version of (2) can be obtained. From this discrete equation, an explicit update expression for the node voltage at time step $n+\frac{1}{2}$, $v_i^{n+\frac{1}{2}}$, can be written in terms of the node voltage, $v_i^{n-\frac{1}{2}}$, and branch currents, $i_{i,p}^n$, at previous time instants as

$$v_i^{n+\frac{1}{2}} = \frac{1-\frac{G_{ii}\Delta t}{2C_{ii}}}{1+\frac{G_{ii}\Delta t}{2C_{ii}}} v_i^{n-\frac{1}{2}} + \frac{\frac{\Delta t}{C_{ii}}}{1+\frac{G_{ii}\Delta t}{2C_{ii}}}(i_{s_i}^n + \sum_{p\,=\,1}^{P_i} i_{i,p}^n). \tag{3}$$

The KVL for the branch between nodes $i$ and $j$ in Figure 15(b) can be written as

$$v_i(t) = R_{ij}i_{ij}(t) + L\frac{di_{ij}(t)}{dt} + v_j(t). \tag{4}$$

Following the same steps followed for node voltages, an explicit update expression for branch currents can be obtained and written as

$$i_{ij}^{n+1} = \frac{1-\frac{R_{ij}\Delta t}{2L_{ij}}}{1+\frac{R_{ij}\Delta t}{2L_{ij}}} i_{ij}^n + \frac{\frac{\Delta t}{L_{ij}}}{1+\frac{R_{ij}\Delta t}{2L_{ij}}}(v_i^{n+\frac{1}{2}} - v_j^{n+\frac{1}{2}}). \tag{5}$$

At each time step, all node voltages are updated first using (3) and all branch currents are updated next using (5). The effect of switching currents are included in (3).

It is to be noted when updating voltages, each node's voltage at time $t$ is updated independently of the voltages of the other nodes at time $t$. The same is true for branch currents. This independence is equivalent to having a diagonal matrix, which can be operated up on with optimal efficiency. This situation is unlike in SPICE-based approaches. In SPICE-based approaches, the node voltage at a particular time instant depends also on the voltages of other nodes at the same instant and on the currents of all branches at the same instant. Therefore, a nonbanded system has to be solved to find node voltages and branch currents.

Because segment sizes are nonuniform, the inductances, $L_{ij}$'s, of the different segments in a single line can be different from each other, and the capacitances to ground, $C_{ii}$'s, from different nodes in the same line can be different from each other. Thus, the time step, $\Delta t$, cannot exceed the Courant time step, $min(\sqrt{L_{ij}C_{ii}})$ [48], for the explicit update expressions, (3) and (5), to be stable, i.e.,

$$\Delta t \leq \min\left(\sqrt{L_{ij}C_{ii}}\right) \tag{6}$$

for all nodes $i$. In (6), the index $j$ refers to the index of the node connected to node $i$ with an inductor.

In the equivalent circuit shown in Figure 13, the number of branches, $N_b$, is linearly proportional to $N_n$. When node voltages are updated using (3) and branch currents are updated using (5), the memory complexity of the whole transient simulation is $O(N_n)$, the time complexity is $O(N_n)$ per time step, and the accuracy scales as $O((\Delta t)^2)$. The time complexity of the overall simulation is $O(N_t N_n)$. The number of time steps, $N_t$, depends on the time step, $\Delta t$, (see (6)), and the total simulation time. Because only a direct solver is used, the accuracy and the numerical robustness (i.e., convergence) of the circuit-FDTD method are good.

### 2.4.2 Frequency-Dependent Equivalent Circuit

The update expressions described in Section 2.4.1 can no longer be used when the equivalent circuit seen at a node (or in a branch) is different from the one shown in Figure 15(a) (Figure 15(b)). This limitation is inherited from the FDTD method [69]. The update expressions of electric and magnetic fields of the FDTD method are different for different media [69]. This limitation is unlike in a SPICE-based approach, in which the formulation is independent of the circuit topology. When a frequency-dependent model shown in Figure 14 is used instead for power-ground lines, the equivalent circuit at a node (and in a branch) is different from the frequency-independent equivalent circuit of the node shown in Figure 15(a) (Figure 15(b)). The

new equivalent circuits observed at a node and in a branch are shown in Figure 16(a) and Figure 16(b), respectively. In the rest of the section, the update expressions are derived for node voltages and branch currents in the frequency-dependent equivalent circuit model of the on-chip PDN.

The telegrapher's equations [70] for a transmission line running parallel to the $z$-axis can be written as

$$\frac{\partial V(\omega)}{\partial z} = -Z(\omega)I(\omega), \tag{7}$$

and

$$\frac{\partial I(\omega)}{\partial z} = -Y(\omega)V(\omega), \tag{8}$$

where $V(\omega)$ and $I(\omega)$ are the voltage and the current, respectively, at position $z$ and at angular frequency $\omega$; and $Z(\omega)$ and $Y(\omega)$ are the series p.u.l. impedance and the shunt p.u.l. admittance, respectively, at $\omega$. Let $v$ and $i$ represent the voltage and the current, respectively, at position $z$ and at time $t$. Using the circuit-FDTD method, a discrete-time solution of (7) and (8) can be obtained. The update expressions are described in detail in [31] and are only stated here. A typical node in the frequency-dependent equivalent circuit of the PDN with one debye term is shown in Figure 16(a). Each node voltage is updated by solving the following system of equations:

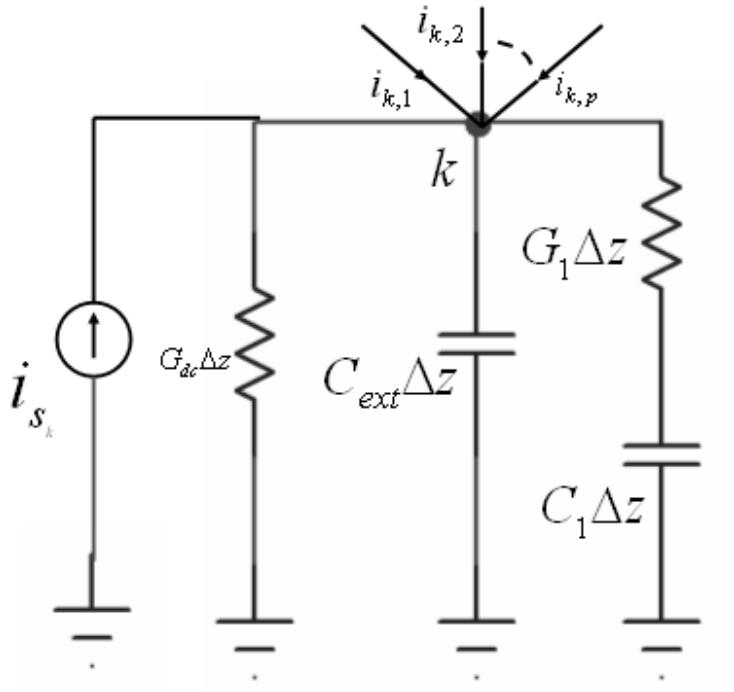
$$\begin{bmatrix} 0 & C_{ext}\Delta z & 0 & 0 \\ 0 & -1 & A^V & 1 \\ 0 & -1 & 1 & Q_1^V \\ 1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} v_k^{n+\frac{1}{2}} \\ b_k^{V,n+\frac{1}{2}} \\ d_k^{V,n+\frac{1}{2}} \\ m_{k,1}^{V,n+\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} RSb_k^V \\ RSd_k^V \\ RSm_{k,1}^V \\ 0 \end{bmatrix}. \tag{9}$$

In (9),

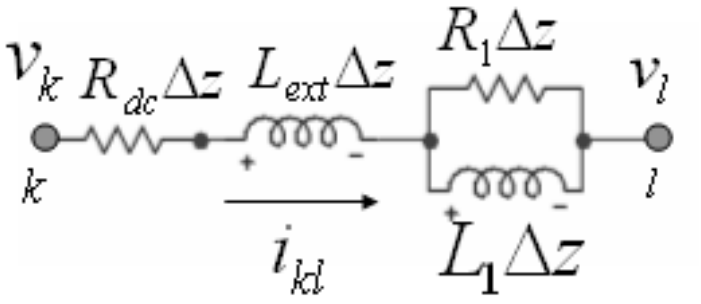
$$A^V = \frac{2C_{ext}}{G_{dc}\Delta t} + 1, \tag{10}$$

$$Q_1^V = \frac{2C_{ext}}{G_{dc}\Delta t} + \frac{C_{ext}}{C_1} + 1, \tag{11}$$

$$RSd_k^V = b_k^{V,n-\frac{1}{2}} + \left(\frac{2C_{ext}}{G_{dc}\Delta t} - 1\right) d_k^{V,n-\frac{1}{2}} - m_{k,1}^{V,n-\frac{1}{2}}, \tag{12}$$




(a) A node




(b) A branch

**Figure 16. A node and a branch in the frequency-dependent equivalent circuit of the on-chip PDN.**

$$RSm_{k,1}^{V} = b_k^{V,n-\frac{1}{2}} - d_k^{V,n-\frac{1}{2}} + \left[\frac{2C_{ext}}{G_{dc}\Delta t} - \frac{C_{ext}}{C_1} - 1\right] m_{k,1}^{V,n-\frac{1}{2}}, \tag{13}$$

and

$$RSb_k^{V} = C_{ext}\Delta z b_k^{V,n-\frac{1}{2}} + \Delta t \Delta i_k^n, \tag{14}$$

where

$$\Delta i_k^n = i_{k-\frac{1}{2}} - i_{k+\frac{1}{2}} + i_{s_k}^n + \sum_{j=1}^{p} i_{k,j}^n. \tag{15}$$

In (15), $i_{s_k}^n$ is the current entering node $k$ as a result of the switching source, $i_{k,i}^n$ is the current entering node $k$ from the $i$th branch connected to $k$, and $p$ is the number of branches connected to $k$ from other lines. It can be noticed from (9) that a small matrix is solved for updating each node voltage. The size of this matrix depends on the number of Debye terms used for a segment, and this number is usually small ($<$ 1-4) and independent of $N_n$. It can also noticed that the new node voltage update expression (9) is different from the previous node voltage update expression (3). Some extra cost is incurred in using the former expression over the latter.

Following a similar procedure for branch currents, the update expressions for branch currents can also be derived. The new update expressions for node voltages and branch currents have the same advantages as those of the update expressions (3) and (5).

## 2.5 DC Simulation

DC simulation is used to set initial conditions for node voltages and branch currents before a transient simulation can be began. As all circuit simulators finally solve ordinary differential equations (ODEs) for the transient simulation, for proper transient results, initial conditions are important. DC simulation involves finding these conditions. If the circuit contains any DC sources (voltage and/or current), then a DC simulation is necessary.

In an on-chip power grid, power lines are connected to each other and also to

power C4 bumps (see Figure 5), which are modeled as ideal DC voltage sources. The same is true for ground lines. Therefore, a DC simulation is needed.

Finding the initial conditions can be easy in some cases. For example, when a circuit contains only DC voltage sources and has no conducting path to ground. In such a case, all nodes in the circuit will have the same voltage as the voltage of the source and all branches will have zero current. Such a case is true in the equivalent circuit proposed in [39], [40], [45]. Therefore, initial conditions were known without running a DC simulation.

In [41], [31], however, the initial conditions are hard to find, requiring a DC simulation. In [41], [31], a new equivalent circuit (proposed as part of modeling the silicon substrate loss) has a conductance to ground (see Figure 13 or Figure 14). Conductance to ground provides a conducting path to ground. When such a circuit is connected to a DC voltage source, the DC current in the power grid is nonzero. As a result, initial conditions cannot be found so easily as they were in [39], [40], [45]. Therefore, in [41], [31], DC simulation was run. For this purpose, the circuit-FDTD method was employed.

Because of the presence of the conductance to ground term, $G_{dc}$, from each node of the on-chip PDN, the voltage applied at a power (or a ground) bump produces a spatial variation of the voltages in the PDN. The DC voltage of a node in the PDN is also the value of its step response at time $t = \infty$. In practice, the step response settles to a value close to the DC operating point by the settling time [71] of the step response. The step response can be simulated through the explicit formulation presented in Section 2.4. The DC simulation can be stopped after step responses at all nodes have settled to a desired tolerance.

When the DC simulation is stopped before step responses at all nodes have settled to a desired tolerance, the circuit would not be in a steady state when the transient simulation starts. This situation happens if the simulation is stopped based on the

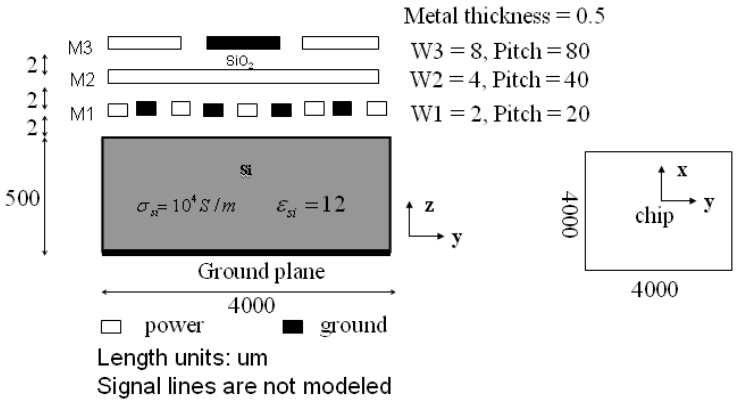



**Figure 17. Cross section of an interdigitated power grid.**

information gathered from only a couple of nodes and not all nodes. When the transient simulation is performed on a circuit with an unsettled step response, the node voltages computed during the transient simulation would have contributions from both the switching source and the step input (because of the power/ground supply bumps). The consequences of this mixed contribution are twofold: 1) the PSN might be erroneous and 2) there might be fluctuations in the voltage of a node even before the effect of a switching source could be felt at the node. Because the effect of the switching source felt at a node is constrained by the speed of light in the medium, the second consequence might appear to be a violation of the causality in the transient simulation. Therefore, for a correct DC analysis using the circuit-FDTD method, it has to be ensured that the step responses at all the nodes are well settled, which requires that the DC simulation be run for a long time.

## 2.6 Results

In this section, the PSN profile in an on-chip PDN with nonuniform line spacing is presented. The effect of the premature termination of the DC simulation on the transient results is illustrated when presenting the results of the irregular PDN. After analyzing the irregular PDN, the effect of the frequency-dependent equivalent circuit on the PSN is shown.

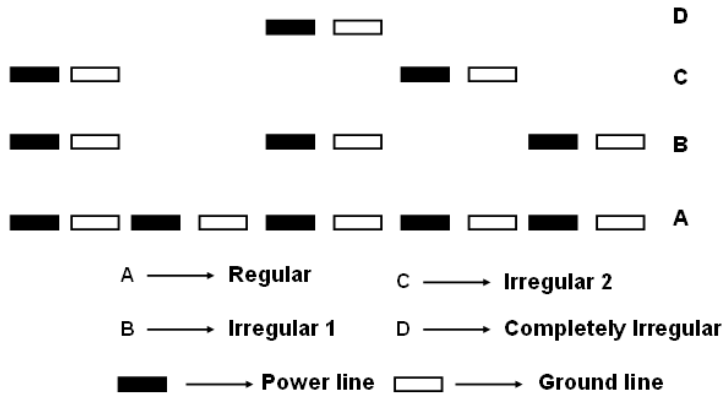



**Figure 18. Irregular arrangement of lines in M1.**

The setup for all tests in this section is a three-metal layer chip shown in Figure 5 and described in Figure 17. Irregularities in the PDN are introduced by removing some power-ground line pairs in M1, as shown in Figure 18. The number of nodes in the PDN is 181,503. For the discussions in this section, it is useful to describe the geometry of the chip with respect to the cartesian coordinate system. Let the chip be placed in the positive quadrant with its left bottom corner at the origin. The lines in M1 and M3 run parallel to the $x$-axis, while the lines in M2 run parallel to the $y$-axis. A linear current source, described in Section 2.3, is used to model the switching circuits. The switching source is a periodic triangular current pulse stream (rise time = 10 ps, fall time = 20 ps, no delay time, periodicity = 200 ps, and peak current = 150 mA) and is applied at (x = 2 mm, y = 2.4 mm) starting from $t = 0$. The simulation is carried out for 0.3 ns. The differential voltages in M1 along an imaginary line x = 2 mm are computed. Each frequency-independent simulation took 5 hours for completion, and each frequency-dependent simulation took 9 hours for completion. The DC simulation in either kind of simulation took approximately 80% of the total simulation time. As can be noticed, majority of the runtime is taken by the DC simulation. This situation is because of employing a circuit-FDTD method

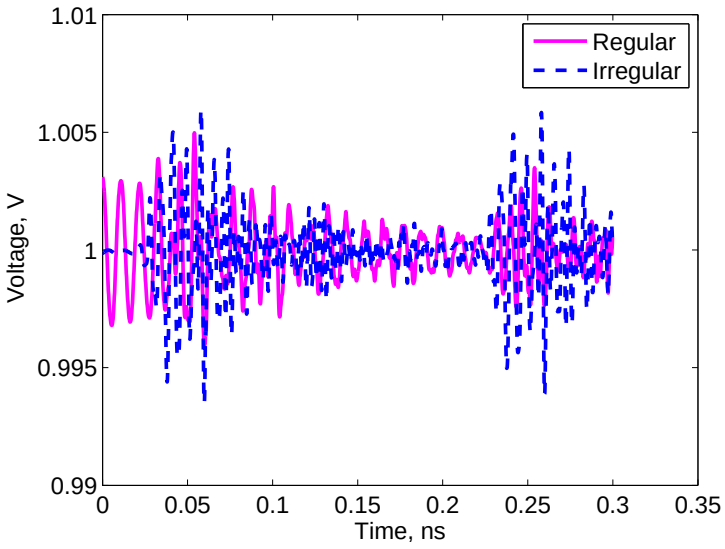
for the DC simulation.

### 2.6.1 Effect of Circuit-FDTD Method-Enabled DC Simulation on PSN

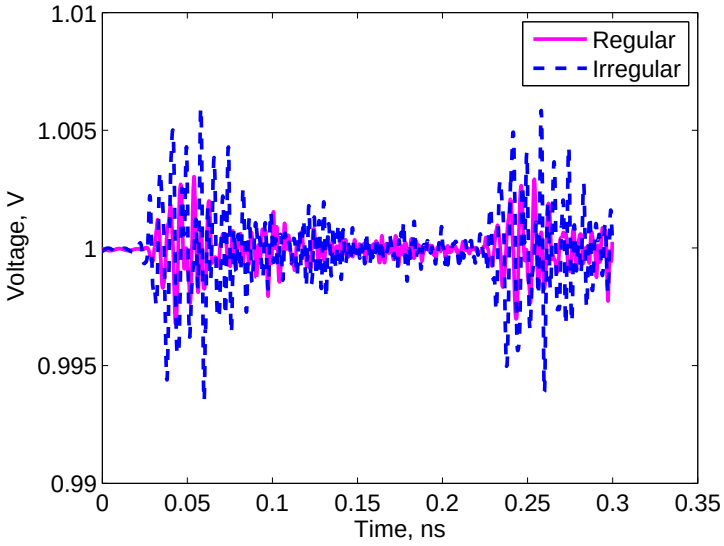
To study the effect of the nonuniform line spacing in the PDN on the PSN, the PSN is computed for the Irregular 1 (See Figure 18) type spacing of the PDN and is compared with the PSN from the Regular (See Figure 18) type spacing of the PDN. The frequency-dependent equivalent circuit (See Figure 14) is used for this test. First, the effect of the premature termination of the DC simulation on the transient results is illustrated. In Figure 19(a), the transient voltage at a node 0.72 mm from the switching source (in a direction perpendicular to the lines) is shown. From Figure 19(a), some oscillations can be noticed near $t = 0$. Because these oscillations are present even before the effect of the switching source can be felt at this location (which is 0.72 mm away from the switching source), these oscillations are not because of the switching source. Hence, these oscillations are spurious. From Figures 19(a-b), it can be noticed that the peak voltage (at time $t = 0.05$ ns) for a regular PDN with the spurious oscillations could be different from the corresponding voltage when these oscillations are minimized. Therefore, the PSN calculations with the result shown in Figure 19(a) can be erroneous. By running the DC simulation for a longer time, these spurious oscillations are minimized significantly (See Figure 19(b)).

### 2.6.2 Demonstration of the Working of the Circuit-FDTD Method in On-Chip PDNs with Nonuniform Power-Ground Line Spacing

After running the DC simulation for a long time, the PSN for the Regular type and the PSN for the Irregular type are computed and compared in Figure 20. From Figure 20, it can be seen that the Irregular type PDN is noisier than the Regular type PDN. The maximum difference between the noise from the irregular PDN and the noise from the regular PDN is 11.04 mV and is found at a distance of 0.24 mm away from source. At this location, the noise from the irregular PDN and the noise from the regular PDN are 28.97 mV and 17.93 mV, respectively. When the irregularity in the

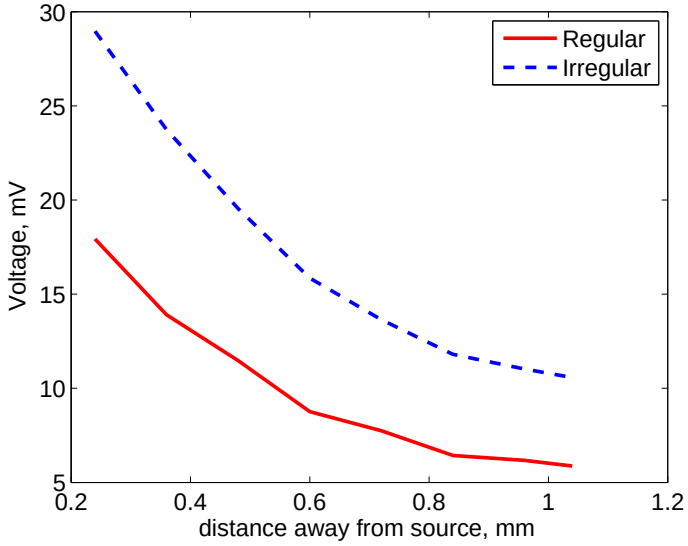



(a) Premature Termination




(b) Mature Termination

**Figure 19. Effect of premature termination of the DC simulation on the transient voltage computed at 0.72 mm away from the switching source.**




(a) Differential noise voltage

**Figure 20. Effect of nonuniform line spacing on the PSN.**

PDN is not modeled and is instead approximated by a regular PDN, the noise at this location is underestimated by 38.1%. The maximum percentage by which the noise is underestimated is 45.5%, and this situation occurs at a distance of 0.84 mm away from the source. The noise from the irregular PDN and the noise from the regular PDN at this location are 11.81 mV and 6.43 mV, respectively. Therefore, from the results presented above, the noise is underestimated by a regular PDN approximation to an irregular PDN by as much as 45.5%. Hence, it may be important to accurately model the irregularity in the line arrangements in the PDN.

### 2.6.3 Effect of Frequency-Dependent Model on PSN

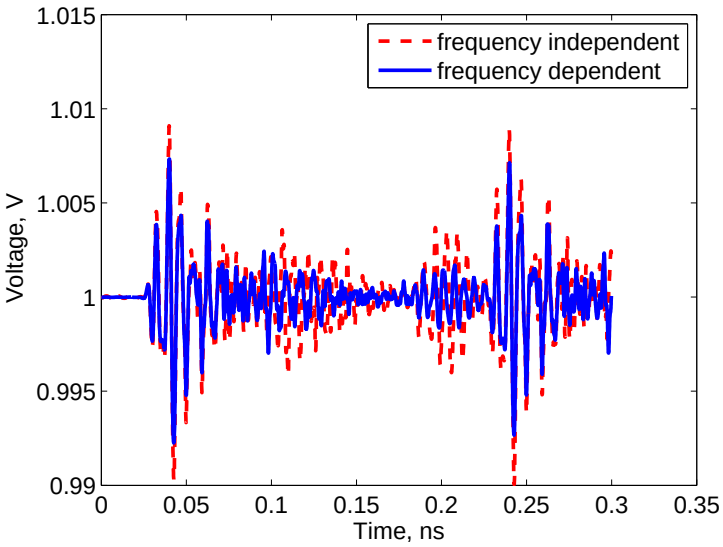
To study the effect of the frequency-dependent variation (which is because of the lossy silicon substrate) of line impedances on the PSN, the PSN is computed for the frequency-dependent equivalent circuit model of the PDN and is compared with that from the corresponding frequency-independent equivalent circuit. For this test, the lines in M1 are made regular (or Regular 1 in Figure 18). In Figure 21, the transient voltage and the PSN in the frequency-dependent equivalent circuit are compared

with those from the frequency-independent equivalent circuit. From Figure 21, it can be noticed that the peak noise decreases as the location gets farther away from the source. The noise from the frequency-independent model is more than the noise from the frequency-dependent model at almost all the locations away from the source. The maximum difference between the noise from the frequency-independent model and the noise from the frequency-dependent model is 2.24 mV and is found at a distance 0.84 mm away from source. At this location, the noise from the frequency-independent model and the noise from the frequency-dependent model are 8.67 mV and 6.43 mV, respectively. The noise is then overestimated by 34.8% with a frequency-independent model. The maximum percentage by which the noise is overestimated is also 34.8%, and this situation occurs at the same location. Thus, from the results presented above, the noise is overestimated by the frequency-independent model by as much as 34.8%. Therefore, it may be important not to ignore the frequency-dependent line parasitics caused by the lossy silicon substrate. However, the total (DC + Transient) simulation time for the frequency-dependent model is almost twice the time taken for the frequency-independent model.

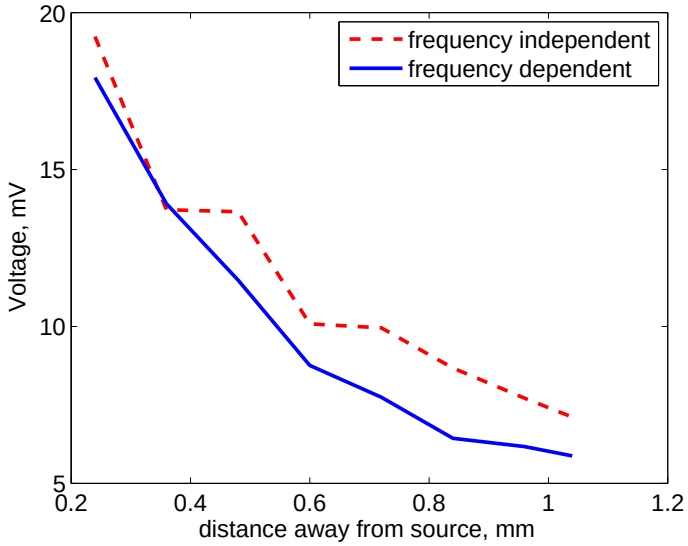
## 2.7 Summary

1. The circuit-FDTD method has been extended to analyze PSN in on-chip PDNs with nonuniform line spacing. *Nonuniform line spacing makes the time step required by the circuit-FDTD method to be different for different nodes in the PDN.* Nonuniform line spacing also increases the loop inductance of power-ground lines and hence may increase the PSN. The increase in the PSN in an on-chip PDN with nonuniform line spacing has been shown.

2. The effect of the frequency-dependent variation of line parasitics on the PSN has been studied. It was shown that when this frequency dependence was ignored, the PSN was overestimated. Therefore, it becomes important to model




(a) 0.72 mm away from source




(b) Differential noise voltage

**Figure 21. Effect of frequency-dependent variation of the line impedances on the PSN in a regular on-chip PDN.**

the frequency-dependence of line parasitics. However, if this dependence was modeled, the total simulation time increases approximately by twofold.

3. Some new drawbacks of the circuit-FDTD method have also been identified when it was used for the DC simulation. It has been observed that when the circuit-FDTD method augmented DC simulation is terminated prematurely, the PSN computed at nodes can be erroneous in amplitude and timing. It was shown that by running the DC simulation for a longer time, the spurious oscillations from the DC simulation can be minimized and therefore the PSN computation can be made accurate. It has also been observed that the circuit-FDTD method-enabled DC simulation takes approximately 80% of the total simulation time. Therefore, accurate and efficient algorithms for the DC simulation are necessary.

# CHAPTER 3

# ACCURATE AND EFFICIENT CIRCUIT-FDTD FORMULATION IN THE PRESENCE OF CROSSOVER CAPACITANCE

## 3.1 Introduction

One of the limiting features of an FDTD-like method for circuits is that the formulation (i.e., the update expressions for voltages and currents) is dependent on the circuit being simulated (see Table 1 in Chapter 1). The equivalent circuits proposed in Chapter 2 did not consider coupling capacitance between lines. When a branch (or coupling) capacitance is considered, not only will the update expressions change, but the linear computational complexity of the update process may be violated. This violation may happen only when solving circuit equations and not when solving Maxwell's equations.

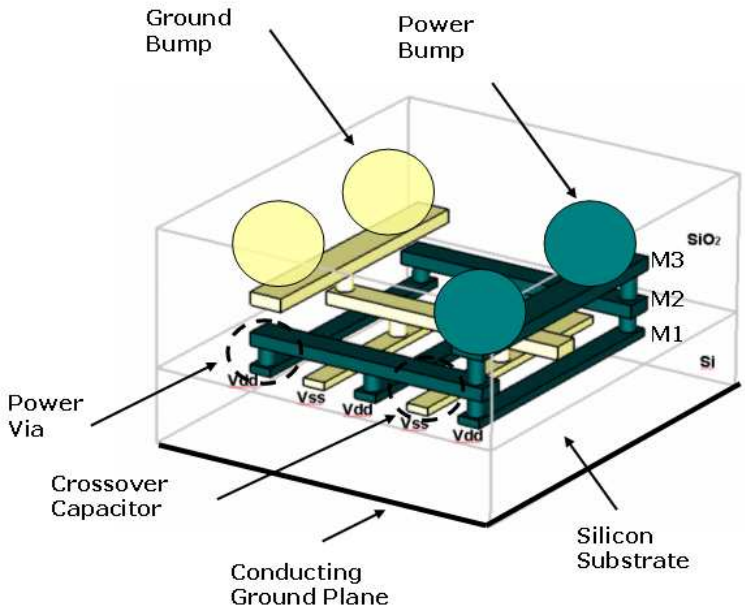



**Figure 22. Crossover capacitance in on-chip PDN.**

Crossover capacitance refers to the overlap capacitance between two lines in adjacent on-chip metal layers (see Figure 22). This capacitance is therefore a branch capacitance. Because the crossover capacitance is present between power-ground lines, it can act as a decoupling capacitance and therefore can affect the PSN. Therefore, it might be important to model this capacitance.

In this chapter, a new formulation for the circuit-FDTD method is proposed in presence of crossover capacitances. Unlike, the prior approaches, the proposed formulation can guarantee linear computational complexity per time step of the transient simulation when these capacitances are considered. The contribution is described at the end of Section 3.2.

## 3.2 Prior Work

FDTD-like formulations including a branch capacitance has been proposed in [48], [39]. However, the formulation in [48] has been shown to be not accurate in [39]. Subsequently, in [39], a more accurate branch capacitance formulation has been proposed. Both the above formulations, however, can violate the linear computational complexity property of an FDTD-like method. In [41], [31], the author has primarily addressed the extraction of capacitance of crossover capacitance.

In this chapter, a new circuit-FDTD-based crossover capacitance formulation is proposed for including the crossover capacitance in both frequency-independent and frequency-dependent equivalent circuits of on-chip PDNs. The proposed formulation guarantees linear computational complexity per time step of the transient simulation unlike [39]. Unlike [41], [31], the proposed formulation includes crossover capacitance in the simulation too. The accuracy and linear computational complexity of the proposed formulation has been demonstrated. *The contribution of this chapter is the proposed formulation.*

The rest of this chapter is organized as follows: In Section 3.3, the crossover

capacitance is explained in detail. In Section 3.4, the new formulation of the circuit-FDTD method with the crossover capacitance is described. In Section 3.5, simulation numerical results showing the effect of the crossover capacitance on the PSN are presented. Finally, in Section 3.6, the conclusions of this chapter are summarized.

## 3.3 Crossover Capacitance

The line-to-line capacitance between lines in adjacent layers, referred to as the crossover capacitance [31] (see Figure 22), comprise both the overlap area capacitance and the fringing capacitance. Because the metal layers below and above a metal layer usually shield the electric flux lines, the crossover capacitance between lines in nonadjacent layers is usually small and therefore is not modeled. The crossover capacitor is present at locations where lines in adjacent layers cross over. This capacitor has a much higher impedance compared to that of the via, even at the highest frequency of operation. This capacitor between power-power (ground-ground) lines in adjacent layers comes in parallel with the low-impedance power (ground) via. Hence, the effect of this capacitance is not felt between power-power lines in adjacent layers. Therefore, the crossover capacitors between power-power (ground-ground) lines in adjacent layers are not modeled. However, the crossover capacitance between power-ground lines in adjacent layers does not have any low-impedance path in parallel. Hence, its effect might be felt between power-ground lines in adjacent layers. Because these capacitances are between power-ground lines, they can act as a decoupling capacitance. The crossover capacitance increases with the decrease in the interlayer thickness and with the increase in the line width and line thickness. See [31] for an extraction procedure for this capacitance.

## 3.4 Formulation with Crossover Capacitance

In this section, the formulation of the circuit-FDTD method with the crossover capacitance is described for the frequency-independent and frequency-dependent equivalent circuits of the on-chip PDN.

### 3.4.1 Frequency-Independent Equivalent Circuit

The node voltage update expression (3) is valid when there is no capacitance between nodes. When a crossover capacitance is included between a power node of a layer and a ground node in the adjacent layer, the node shown in Figure 15(a) would also have a capacitance incident on it. In Figure 23, the modified node $i$ in the PDN with crossover capacitance, $C_{ij}$, from node $j$ is shown. The KCL for the modified node $i$ is given by

$$C_{ii}\frac{dv_i(t)}{dt} + G_{ii}v_i(t) + C_{ij}\frac{d(v_i - v_j)}{dt} = i_{s_i}(t) + \sum_{p\,=\,1}^{P_i} i_{i,p}(t). \tag{16}$$

From (16), an explicit update equation for $v_i^{n+\frac{1}{2}}$ cannot be obtained like in (3). Because $v_j^{n+\frac{1}{2}}$ is also not known, (16) has to be solved simultaneously with the KCL for node $j$. The KCL for node $j$ can be written as

$$C_{jj}\frac{dv_j(t)}{dt} + G_{jj}v_j(t) + C_{ji}\frac{d(v_j - v_i)}{dt} = i_{s_j}(t) + \sum_{p\,=\,1}^{P_j} i_{j,p}(t). \tag{17}$$

Upon approximating the $\frac{d}{dt}$ terms using the central-difference approximation, the continuous-time equations in (16) and (17) can be discretized. The resultant discretized equations can be represented in matrix form as

$$\begin{bmatrix} C_{ii} + C_{ij} + \frac{G_{ii}\Delta t}{2} & -C_{ij} \\ -C_{ij} & C_{jj} + C_{ji} + \frac{G_{jj}\Delta t}{2} \end{bmatrix} \begin{bmatrix} v_i^{n+\frac{1}{2}} \\ v_j^{n+\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} C_{ii} + C_{ij} - \frac{G_{ii}\Delta t}{2} & -C_{ij} \\ -C_{ij} & C_{jj} + C_{ji} - \frac{G_{jj}\Delta t}{2} \end{bmatrix} \begin{bmatrix} v_i^{n-\frac{1}{2}} \\ v_j^{n-\frac{1}{2}} \end{bmatrix} + \begin{bmatrix} \Delta t i_i^e \\ \Delta t i_j^e \end{bmatrix}, \tag{18}$$

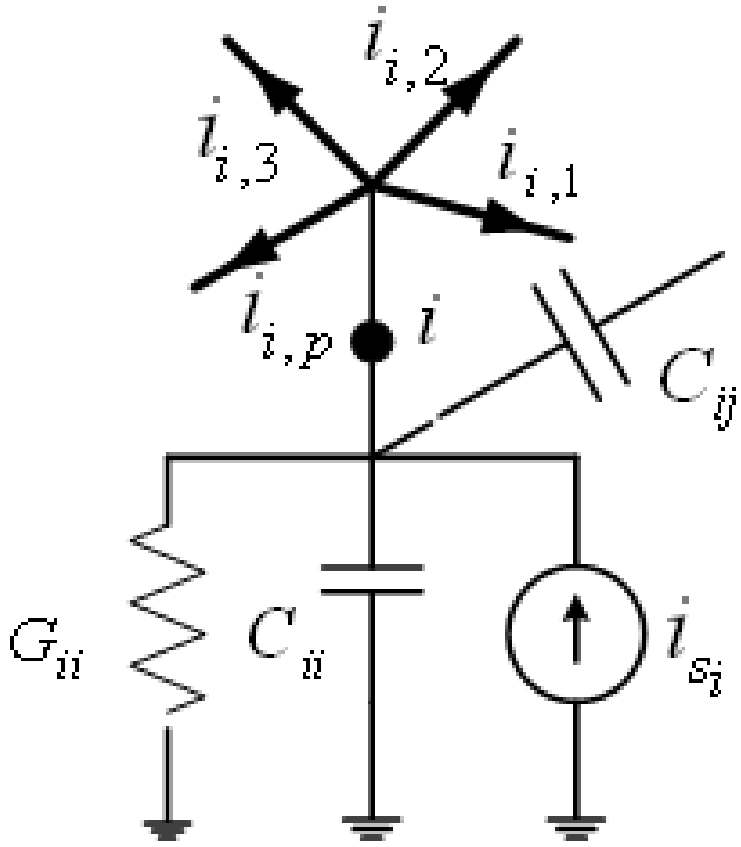



**Figure 23. A node $i$ with a crossover capacitance, $C_{ij}$, from node $j$ in a frequency-independent equivalent circuit of the on-chip PDN.**

where $i_i^e$ is the total current entering node $i$, and $C_{ji} = C_{ij}$. A similar procedure can be extended to the case when more than one capacitor is connected to $i$ from other nodes. In such a case, for every new capacitor, $C_{ik}$, a new row is added to the matrix system shown in (18). The size of this system is equal to the number of such capacitively coupled nodes. Because node $i$ and node $j$ exist in different layers, the maximum number of capacitively coupled nodes would be equal to the number of layers in the PDN. Because the number of metal layers in a PDN is usually a small ($\leq 15$) number, only a small system is solved at each time step. Thus, the linear time complexity and memory complexity per time step of the circuit-FDTD method are not compromised. Moreover, since central differencing is still retained in obtaining (18), the accuracy of the formulation is still $O\left((\Delta t)^2\right)$. Unlike the voltage update equation, the current update equation in (5) would not be changed when a crossover capacitance is present because the branch structure shown in Figure 15(b) is not changed. Therefore, the new formulation presented in this chapter retains the original advantages (see Section 2.4.1 of Chapter 2) of the circuit-FDTD formulation presented in Chapter 2.

### 3.4.2 Frequency-Dependent Equivalent Circuit

With a crossover capacitance, the typical node in the frequency-dependent equivalent circuit will look like that shown in Figure 24. For node $k$, the KCL can be obtained from (8) and can be written as

$$-\Delta I_k = -Y\Delta z V_k - j\omega C_{kq}(V_k - V_q), \tag{19}$$

where $\Delta I_k$ is the net current entering node $k$; and $C_{kq}$ is the capacitance between nodes $k$ and $q$. When the derivatives in (19) are discretized and the resultant expression is simplified, (19) becomes

$$C_{kq} v_k^{n+\frac{1}{2}} + C_{ext}\Delta z b_k^{V,n+\frac{1}{2}} - C_{kq} v_q^{n+\frac{1}{2}} = RSb_{k,q}^V, \tag{20}$$

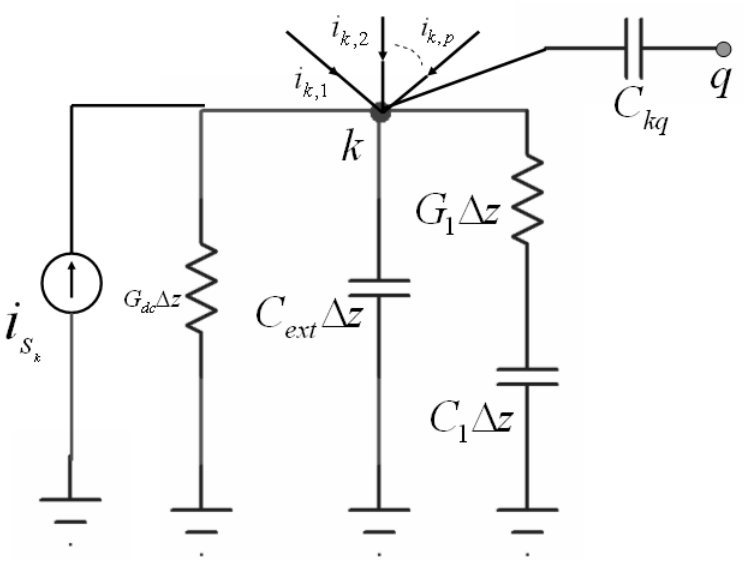



**Figure 24. A node with one debye term and a crossover capacitance, $C_{kq}$, in the frequency-dependent equivalent circuit of the on-chip PDN.**

where

$$RSb^V_{k,q} = C_{kq}v_k^{n-\frac{1}{2}} + C_{ext}\Delta z b_k^{V,n-\frac{1}{2}} - C_{kq}v_q^{n-\frac{1}{2}} + \Delta t \Delta i_k^n. \tag{21}$$

Because $v_k^{n+\frac{1}{2}}$ and $v_q^{n+\frac{1}{2}}$ are simultaneously (i.e., at $n+\frac{1}{2}$) related in (20), $v_k^{n+\frac{1}{2}}$ cannot be explicitly updated using (9) alone. In fact, $v_k^{n+\frac{1}{2}}$ and $v_q^{n+\frac{1}{2}}$ have to be updated simultaneously with their corresponding $b^{V,n+\frac{1}{2}}$, $d^{V,n+\frac{1}{2}}$, and $m_i^{V,n+\frac{1}{2}}$'s. Using (20), performing the same steps that were followed to obtain (9), and repeating the procedure thus far discussed for node $q$ as well, a combined matrix system can be

obtained, as shown in (22).

$$\left[\begin{array}{cccc|cccc} C_{kq} & C_{ext}\Delta z_{l_1} & 0 & 0 & -C_{kq} & 0 & 0 & 0 \\ 0 & -1 & A^V_{l_1} & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & Q^V_{l_1,1} & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline -C_{qk} & 0 & 0 & 0 & C_{qk} & C_{ext}\Delta z_{l_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & A^V_{l_2} & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & Q^V_{l_2,1} \\ 0 & 0 & 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \left[\begin{array}{c} v_k^{n+\frac{1}{2}} \\ b_k^{n+\frac{1}{2}} \\ d_k^{n+\frac{1}{2}} \\ m_{k,1}^{V,n+\frac{1}{2}} \\ \hline v_q^{n+\frac{1}{2}} \\ b_q^{n+\frac{1}{2}} \\ d_q^{n+\frac{1}{2}} \\ m_{q,1}^{V,n+\frac{1}{2}} \end{array}\right] = \left[\begin{array}{c} RSb^V_{k,q} \\ RSd^V_k \\ RSm^V_{k,1} \\ 0 \\ \hline RSb^V_{q,k} \\ RSd^V_q \\ RSm^V_{q,1} \\ 0 \end{array}\right] \tag{22}$$

The subscript $l_1$ and $l_2$ in (22) represent, respectively, the lines on which nodes $k$ and $q$ reside. Also, in (22), $C_{kq} = C_{qk}$. The procedure thus far described can be easily extended to the case when there is more than one term in the debye model of the lines. The arguments for time and memory complexities and for accuracy (of the transient simulation) with the crossover capacitance in the frequency-dependent equivalent circuit are the same as those discussed in the frequency-independent equivalent circuit and therefore are not discussed here.

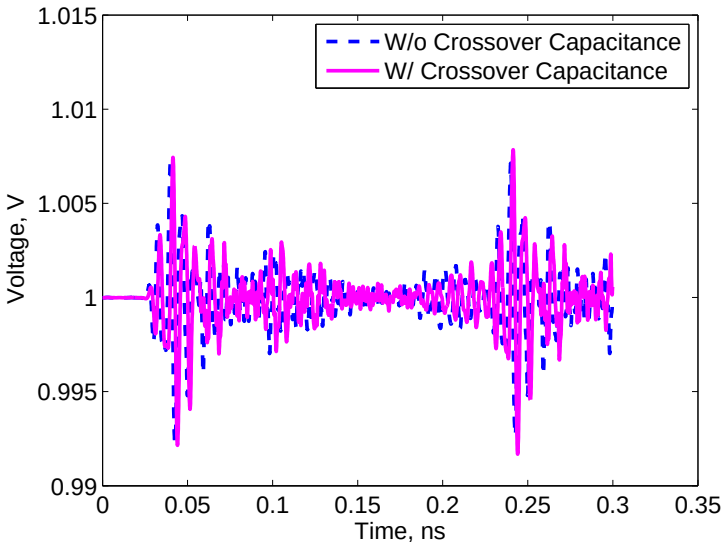
## 3.5 Results

In this section, simulation results showing the effect of the crossover capacitance on the PSN are presented. The PSN is computed for a regular PDN of type Regular 1 with and without the crossover capacitance. Though the accurate expression for the crossover capacitance is derived in [31], a simple parallel-plate capacitance value is chosen for the crossover capacitance. The parallel plate capacitance between lines with widths $w_1$ and $w_2$ and with distance between them, $d$, is given by $C_{parallel} = \epsilon_{SiO_2} w_1 w_2 / d$. This capacitance is 0.41 fF between lines in M1 and M2 and is 1.63 fF between lines in M2 and M3. In Figure 25, the PSNs in the frequency-dependent

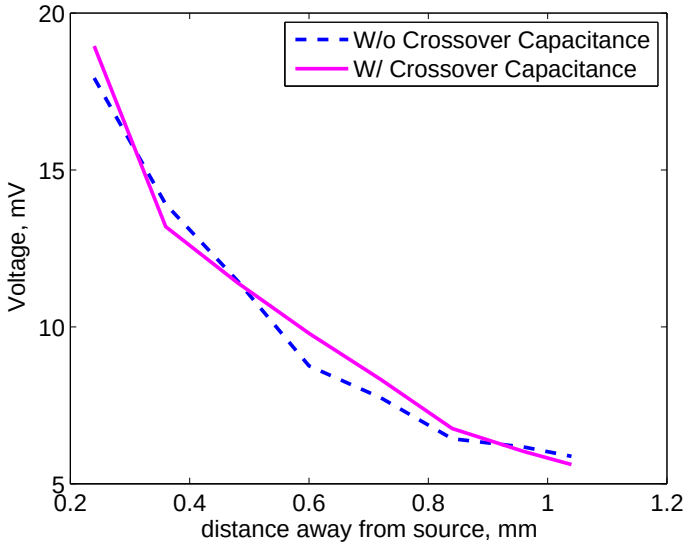
model of the regular PDN computed with and without the crossover capacitance are compared. From Figure 25(b), it can be observed that the PSN (computed) with the crossover capacitance is different, though not to a great extent, from that without the crossover capacitance; it can be observed that the crossover capacitance does not necessarily result in reduced noise amplitudes, an expected outcome considering that the crossover capacitance can act as a decoupling capacitance (as it is between a power line and a ground line). Though the difference in noise with and without the crossover capacitance is not significant, it is still beneficial to model this capacitance, as the noise calculations are more accurate with the crossover capacitance.

## 3.6 Summary

The equivalent circuit of the on-chip PDN has been updated by including the inter-layer line-to-line capacitance, known as the crossover capacitance, between power-ground lines in adjacent metal layers. The formulation of the circuit-FDTD method has been modified to include this capacitance in both the frequency-independent and frequency-dependent equivalent circuits of the on-chip PDN. The new formulation, unlike the prior approaches, guarantees linear computational complexity per time step of the transient simulation. Simulation results showing the effect of this capacitance on the PSN have been shown. It has been found that the crossover capacitance affects the PSN, but not significantly.




(a) Transient voltage at 0.72 mm away from the source




(b) Differential noise voltage with increasing distance away from the source

**Figure 25. Comparison of the differential voltage and the differential noise obtained with and without the crossover capacitance in a frequency-dependent model of a regular on-chip PDN.**

# CHAPTER 4
# ACCURATE AND EFFICIENT DC SIMULATION

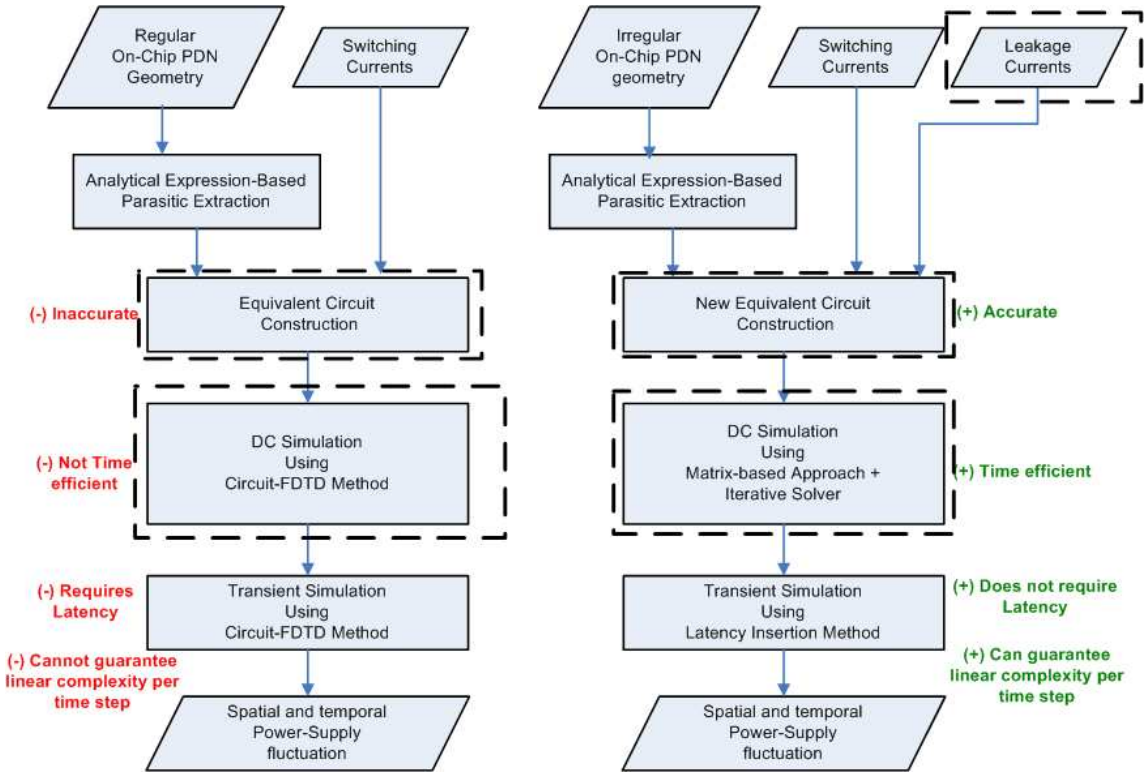
## 4.1 Introduction

In Chapter 2, it was identified that the circuit-FDTD method-enabled DC simulation can become inaccurate in some cases and time inefficient in all cases. Therefore, there is need for a more accurate and efficient DC simulator. In this chapter, it is proposed to perform DC simulation using a SPICE-based approach (see Section 1.2.2.1) augmented with an iterative solver. The proposed approach has been demonstrated to have better run times than the circuit-FDTD method. The focus of this chapter is also highlighted in Figure 26. *The contribution of this chapter is the demonstration of the time inefficiency of the circuit-FDTD method for DC simulation.*

## 4.2 Prior Work

In [41], [31], the presence of a conductance term to ground, $G_{dc}$, (see Figures 13 and 14), in the on-chip equivalent circuit setup a conducting path to ground. Existence of such a path made the initial conditions hard to find without a DC simulation. Circuit-FDTD method was used for this simulation. This method for the DC simulation was shown (in Chapter 2) to be requiring majority (almost 80%) of the total simulation (i.e., DC + transient) time. Therefore, it is important to find if this situation can be avoided or at least be improved.

Apart from the time inefficiency of the DC simulation in [41], [31], there are two more shortcomings to [41], [31]. These shortcomings concern the equivalent circuit employed for DC simulation. First, DC simulation is required in on-chip PDNs because of the presence of a DC current source. This current source can be an approximation to the average DC current consumed during switching or to the leakage current consumed during both switching and nonswitching times. However, in [41], [31], no




(a) Prior approach [45], [39], [40], [41], [31].

(b) Proposed approach in this dissertation.

**Figure 26. Comparison of the prior and proposed approach in FDTD-based circuit simulation of PSN in on-chip power grids. The focus of this chapter is the feature in the figure marked within the dashed rectangle.**

DC current sources were included. Second, the conductance to ground term, $G_{dc}$, in the on-chip PDN equivalent circuit cannot be physical (explained in detail later in this chapter).

When circuit-FDTD method is employed for on-chip PDN DC simulation with a DC current source, then the run time significantly worsens. The run time worsens because the step responses at nodes now will have a much higher magnitude of fluctuation (than what it is was without a DC current source) and consequently take a much longer time to settle. The increased magnitude of fluctuation has to do with injecting a current suddenly (recall in a circuit-FDTD method, DC sources are only modeled as transient step sources) in a circuit that has nonzero inductance. Increasing the rise times of the step current sources or decreasing the inductance are options that can be explored to improve the runtime. Another option worth considering is not to perform a transient simulation (which is what was essentially done when circuit-FDTD method was employed) for a DC simulation.

Traditionally, this is how DC simulation is performed. For a DC simulation, inductors are shorted and capacitors are left open, leaving only a resistive circuit with DC sources.

Traditionally, DC simulation is performed through a SPICE-based approach. A direct solver would be preferred if computational complexity can be ensured either to be linear or close to being linear. However, in general, using a direct solver makes the computational complexity dependent on the numbering of nodes.

In the absence of a direct solver, there are two choices for the solver in the context of a DC simulation. First, an iterative solver can be used, see [35], [36]. The drawback of using this solver is that the convergence may be slow. Accuracy is compromised a little, but is observed to be tolerable. Considering the sparse nature of the matrix, an iterative solution may be preferred. Second, a statistical solver can be used, see [34]. A statistical solver-based on random walks for the power grid DC simulation has

been shown to be computationally efficient [34]. In this chapter, DC simulation is performed using a SPICE-based approach augmented with an iterative solver.

In this chapter, it is shown that the $G_{dc}$ term is not present in on-chip PDNs, leaving no conduction path to ground in the on-chip PDN equivalent circuit. Its absence is observed to make the spatial variation of the DC node voltages trivial when only the power-supply voltages were applied. When a nonzero average current is also flowing into/out of a node, the spatial variation of the DC node voltages can be nontrivial, even with no conducting path to ground. This fact is used to include the leakage current, which is predominant in chips nowadays, as the nonzero average current producing the DC IR drops. It is shown that the leakage current can produce a significant DC IR drop in on-chip PDNs. The DC node voltages are efficiently computed by solving the MNA matrix using an iterative solver. This new DC IR-drop simulator is shown to be more accurate and computationally more efficient than the circuit-FDTD method. No significant convergence problems have been observed. *The contribution of this chapter is the demonstration of the time inefficiency of the circuit-FDTD method for DC simulation.*

The rest of this chapter is organized as follows: In Section 4.3, the reason for dropping the $G_{dc}$ term is explained, and the consequences of not having $G_{dc}$ term on the DC simulation are explained. In Section 4.4, the effect of the leakage current on the DC IR-drop and the circuit model for the leakage current are discussed. In Section 4.5, the new technique for the DC analysis is briefly described. In Section 4.6, a sample result showing the effect of the leakage current on the DC IR drop is presented, and the memory and the computational time of the new technique are reported. Finally, in Section 4.7, the conclusions of this section are summarized.

## 4.3 $G_{dc}$ term in the On-Chip PDN Equivalent Circuit

The term $G_{dc}$ stands for the conductance to ground at DC. This is also the G term in the RLGC model of a transmission line over a dielectric with nonzero conductivity. This term is present in the transmission-line models of signal lines residing in the package or in the PCB where the dielectric might have a nonzero finite conductivity. However, in on-chip power grids, power and ground lines are immersed in silicon dioxide (see Figure 36), which has zero conductivity, and are not in direct contact with the silicon substrate, which has a nonzero conductivity. Therefore, the transmission-line-based equivalent circuits of the power/ground lines in any on-chip metal layer do not have a $G_{dc}$ term. However, they can have a conductance term in series with a capacitance, as shown in Figure 14, to model the dielectric loss in the dielectric (silicon dioxide + silicon substrate).

When all the $G_{dc}$ terms are set to zero in the equivalent circuit (see Figure 13 or Figure 14) of the on-chip PDN, there would be no passive conducting path from any node to the ground (i.e., system ground). In such a circuit, if the switching current were absent and only the power-ground supply voltages were applied (at the C4 locations), then there would not be any DC IR drop in the PDN. Therefore, all the power nodes would have a DC voltage of $V_{dd}$, and all the ground nodes would have a DC voltage of $V_{ss}$. This setup makes the computation of the DC node voltages trivial. This situation was the reason why [39], [40], [45] did not run a DC simulation. Without the $G_{dc}$ terms, the equivalent circuit of the on-chip PDN for DC analysis is shown in Figure 27. The leakage current sources in Figure 27 are described next.

## 4.4 Leakage Current and IR drop

When a nonzero average current is flowing into/out of a node in a resistive circuit with no conducting path to ground, there would be voltage drop in the circuit. Usually, the DC analysis is performed with an average current, calculated from the total switching

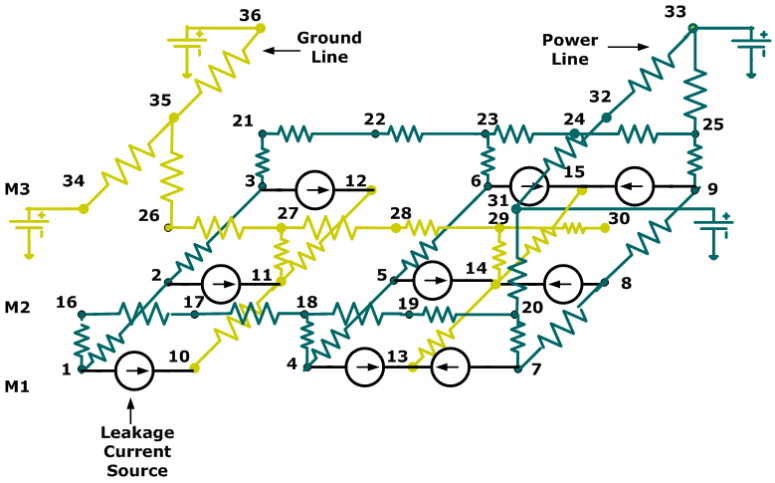



**Figure 27. The on-chip PDN equivalent circuit used for DC analysis.**

power dissipated and the power-supply voltage. Such average-current estimates are used in the early mode analysis (see [27]) of power grids to decide locations of C4 bumps, nominal pitches of lines, and widths of lines. Because the switching currents are actually transients, the average IR drop induced by them would be comparable to the DC IR drop calculated from the average current estimates only if most of the circuits are switching and that too for a long time, which has a low probability of occurence typically. Therefore, the DC IR drops computed with an average value for switching currents are most probably worst-case values.

In future technology nodes, the power dissipation from the leakage current is predicted to be more than the power dissipation from the switching current [19]. The leakage current can have a significant nonzero average current and hence can cause DC IR drops. Because the leakage current is present in the circuit whether or not the circuits switch, the leakage current is present at all times. Unless otherwise designed for reduced leakage current, all circuits leak. Therefore, it the leakage current is present in most parts of the chip. *Because the leakage current is present at all times and in most parts of the chip, the IR drop resulting from it is also going to be present at all times and in most parts of the chip.* This situation is opposite to the one caused by the switching current. Because the leakage power is expected to be more

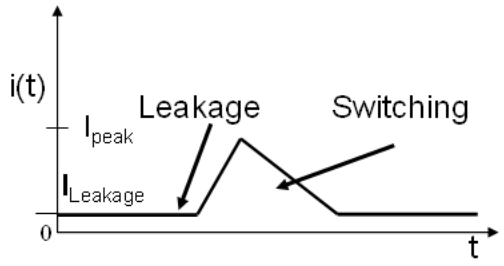



**Figure 28. Comparison of the leakage current and the switching current with time.**

than the switching power, the DC IR drop contribution from the former may be more significant than that from the latter.

The leakage power dissipated is given by

$$P_{Leakage} = V_{dd} I_{Leakage}, \tag{23}$$

where $P_{Leakage}$ is the average leakage power dissipated, and $I_{Leakage}$ is the average leakage current flowing out of the power supply voltage, $V_{dd}$. Based on (23), the leakage current is modeled as a DC current source (see Figure 28) whose magnitude is computed using (23) knowing $P_{Leakage}$ and $V_{dd}$. This current source is placed between a power node and the ground node closest to the power node. In Figure 27, some leakage current sources in the on-chip PDN are shown.

## 4.5 Efficient DC Analysis

The DC analysis can be performed in the new equivalent circuit (the one having no $G_{dc}$, but having the leakage current) using the circuit-FDTD method. However, with a step current source (for the leakage current), the peak-to-peak voltage oscillations are going to be more than those with only a step voltage source (for the C4 power-supply bumps). These high-amplitude oscillations increase the time it takes for these oscillations to be smaller than the desired tolerance. As a result, the total time taken to complete the DC simulation (using the circuit-FDTD method) increases

significantly, as will be reported in Section 4.6. Therefore, a new simulation technique for the DC analysis is needed that is more accurate and computationally more efficient than the circuit-FDTD method.

Because the circuit-FDTD method restricts the time step, the DC simulation times are usually longer. Thus, the DC analysis is performed by solving the static on-chip PDN equivalent circuit: only the resistors, power and ground DC voltage supplies, and the leakage current sources are kept; the inductors are shorted, and the capacitors are made open circuits. The resulting resistive circuit is casted using the MNA, as is being done in SPICE [3]. However, unlike SPICE, the sparse system with a symmetric coefficient matrix is solved using an iterative solver. In this work, the transpose-free quasi-minimal residual algorithm [72] is used as the iterative solver. The computational complexity of the matrix-vector product is $O(N_n)$, and the memory complexity of the solution is $O(N_n)$. The computational time complexity depends on the number of iterations, which depends on the condition number of the coefficient matrix. A reasonably good convergence has been observed for all problem run thus far.

## 4.6 Results

In this section, a simulation result showing the effect of the leakage current on the DC IR-drop is presented and the performance of the new DC simulator (MNA + iterative solver) is compared with that of the circuit-FDTD method.

### 4.6.1 Effect of Leakage Current on DC IR Drops

The test setup for the results in this section is same as the one described in Section 3.1 with the following changes: 1) The power-ground supply C4 bumps are arranged as shown in Figure 29(a), and 2) a leakage power of 125 mW mm$^{-2}$ is uniformly distributed in M1. In M1 (4 mm $\times$ 4 mm area), the total leakage power dissipated is 2 W. Assuming a 1 V supply voltage, this leakage power corresponds to 2 A of

leakage current, which is distributed evenly among 181,503 nodes (49 uA per node). The leakage power value is chosen such that it is 50% of the total power dissipated, with a 250 mW $mm^{-2}$ total power (switching + leakage) dissipation density.

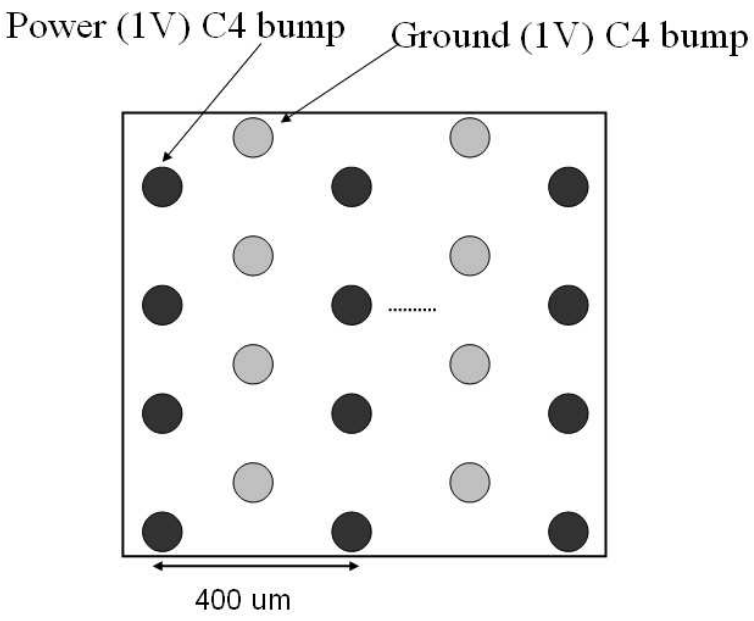
In Figure 29, the distribution of the DC voltages in one-fourth of the area of M1 is shown. There was a maximum DC IR drop of 3.8 mV from the ideal value of 1 V. Though this drop is small ($\leq$ 1% of $V_{dd}$), it would be comparable to the magnitude of the average voltage drop if the 2 W were instead dissipated as a switching power. Therefore, it is important to include leakage currents in PSN computation.

### 4.6.2 Circuit-FDTD Method Vs. Proposed Method: Performance Comparison

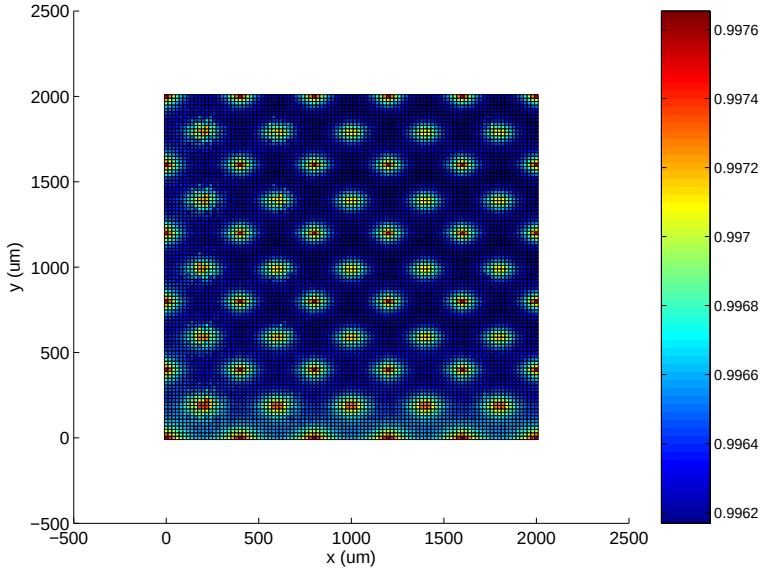
The DC simulation for the sample problem (the one with 181K nodes) was performed using both the circuit-FDTD method and the new method (MNA + Iterative Solver). Both the methods are accurate, as they capture the C4 foot print in the DC voltage distribution. The accuracy of these methods have been verified against HSPICE for small problems. The memory requirements for these methods are 47 MB and 74 MB, respectively. The computational time requirements for these methods are 39 hours and 12 minutes, respectively. Therefore, the new method for the DC simulation is much more efficient than the circuit-FDTD method. This problem was also tried in HSPICE but could not be completed because of high memory requirements ($>$ 1 GB).

## 4.7 Summary

In this section, it has been shown that the conductance to ground term, $G_{dc}$, would not be present in the equivalent circuits of power/ground lines. It has been described that the absence of this term results in zero DC IR drop in the DC node voltages when there are no DC current sources present. It has been described and shown that the leakage current can produce DC IR drop, even without the $G_{dc}$ term. A constant current source model for the leakage current has been proposed. The DC analysis




(a) Arrangement of the power-ground supply bumps in M3.




(b) DC voltage distribution in one-fourth area of M1

**Figure 29. Spatial distribution of the DC node voltages in one-fourth area of M1 due to a uniform distribution of the leakage current in M1. Leakage power density is 125 mW mm$^{-2}$, Area of M1 is 4 mm $\times$ 4 mm. Maximum DC IR-drop is 3.8 mV.**

has been performed more time efficiently than the circuit-FDTD method, by solving an MNA system using an iterative method.

# CHAPTER 5

# SIMULATION OF POWER-SUPPLY NOISE IN IRREGULAR ON-CHIP PDNS USING CIRCUIT-FDTD METHOD

## 5.1 Introduction

In Chapter 2, the circuit-FDTD method was extended to compute the PSN in on-chip PDNs with nonuniform line spacing. However, the power/ground lines analyzed had uniform cross section along its length and ran from one side of the chip to the other without being discontinuous. In this chapter, the circuit-FDTD method is extended to compute the PSN in on-chip PDNs where the lines are discontinuous and have nonuniform cross sections (see Figure 6). When lines have nonuniform cross section, the parasitics of the line vary along the line. This variation alters the Courant time step (or the maximum time step) of the circuit-FDTD method. When lines are discontinuous, finding the locations of vias and crossover capacitances becomes difficult. For the transient simulation, the new formulation (w/ crossover capacitance) proposed in Chapter 3 is employed. For the DC simulation, the new method described in Chapter 4 is employed. *The contribution of this chapter is the application of circuit-FDTD method for transient PSN simulation in irregular power grid geometries*[1].

The rest of this chapter is organized as follows: In Section 5.2, the changes to the simulation when analyzing irregular on-chip PDNs are presented. In Section 5.3, sample PSN results in irregular PDNs demonstrating the accuracy of the simulation are presented. Finally, in Section 5.4, conclusions of this chapter are summarized.

[1]**S. N. Lalgudi**, M. Swaminathan, and Y. Kretchmer, "Simulation of simultaneous switching noise in on-chip power distribution networks of FPGAs," *IEEE 14th Topical Meeting on Electrical Performance of Electronic Packaging*, Oct. 2005, pp. 319-322.

## 5.2 Changes to the Simulation

### Changes to Circuit-FDTD Method

When a line has nonuniform cross section (see Figure 6), the line is modeled by concatenating several uniform sections. As the line width changes, the parasitic line inductance and capacitance also change. Therefore, the Courant time step at each node is affected and is still given by (6).

### Changes to Geometry Processing

When a line is discontinuous (see Figure 6), a new problem related to finding the locations of vias and crossover capacitors arises. Vias are present between power-power (or ground-ground) lines in adjacent layers, and crossover capacitances are present between power-ground lines in adjacent layers. When all lines are continuous (see Figure 30(a)) and run from one side of the chip to the other, then each power (ground) line in a layer would contribute to as many vias as there are power (ground) lines in the adjacent layer and to as many crossover capacitances as there are ground (power) lines in the adjacent layer. If the lines in each metal layer have uniform line spacing, then the locations of all vias and crossover capacitors contributed by a line can be computed knowing the pitch of the lines in the adjacent metal layer. Therefore, just by knowing the pitch of the lines, the number of vias and crossover capacitors, the locations of the vias and the crossover capacitors can be determined. This process can be accomplished in $O(N_l)$ computational time, where $N_l$ is the total number of lines in the PDN.

When lines are discontinuous, the number of vias and crossover capacitors and their locations (see Figure 30) are not as they were when lines were continuous. The locations of vias (crossover capacitances) in discontinuous lines are determined by projecting a line in a metal layer on the adjacent layer and by finding the intersection (if any) of each line segment with the other line segments. The problem reduces to computing the intersection of $N_l$ line segments. The computational time required for

finding all the line-segment intersections scales as $O(N_l^2)$. This $O(N_l^2)$ time complexity can a bottleneck if $N_l$ is linearly proportional to $N_n$. This situation can happen in the PDN geometry at the post-layout design stage where the PDN can be highly irregular.

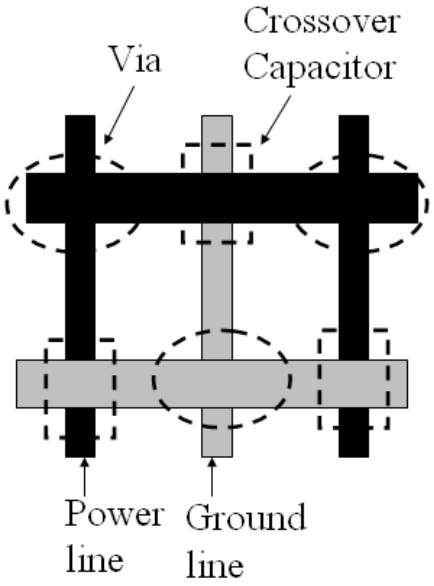
## 5.3 Results

In this section, the PSNs are computed in on-chip PDNs in which lines are discontinuous and in which lines have nonuniform cross sections. The DC analysis is carried out using the new technique described in Chapter 4.

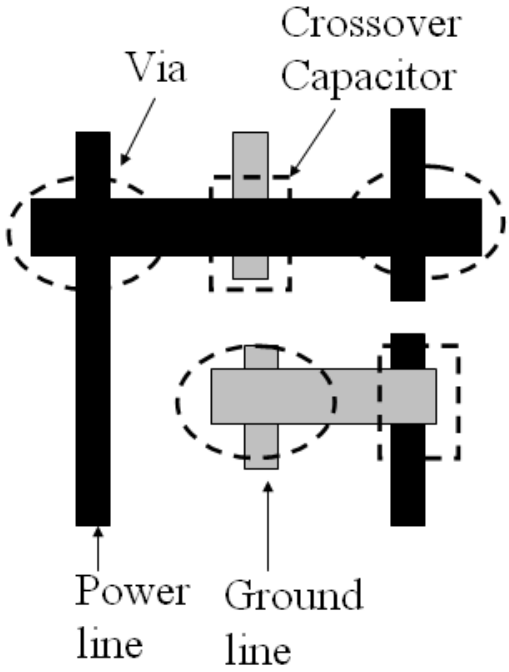
The test setup is same as the one considered in Section 3.1 with the following changes: The leakage power specifications described in Section 4.6 are followed and a switching power of 2 W is dissipated in 1 $mm^2$ area of M1. In Figure 31, the switching and leakage current sources and the output locations are described.

### 5.3.1 Effect of Nonuniform Cross-Section of Lines on PSN

To observe the PSN profile in an on-chip PDN with nonuniform cross section, the following test was done. Each line in M1 is broken into two sections. One section is made to have the same width as before. The other section is made to have twice the width. The differential voltages at all power nodes in M1 are computed. In Figure 32, the voltages at $t = 35$ ps computed with and without the different widths are compared. From both Figure32(a) and Figure 32(b), it can be observed that the disturbance travels faster along the length of the lines ($x$-direction) than it does along the width of the lines ($y$-direction). The extra time taken in the latter is because of the extra distance (M1 to M2 to M1) the signals have to travel (see [31] for a detailed description). Capturing this difference in the propagation times qualitatively verifies the accuracy of the simulation. The voltage distribution in Figure 32(a) is different from that in Figure 32(b). The minimum differential voltage for the nonuniform case is 0.9448 V and occurs at (x = 2.36 mm, y = 1.98 mm). The voltage at the same




(a) Continuous lines.




(b) Discontinuous lines.

**Figure 30. Comparison of the via and crossover capacitor locations in on-chip PDNs with and without continuous lines.**

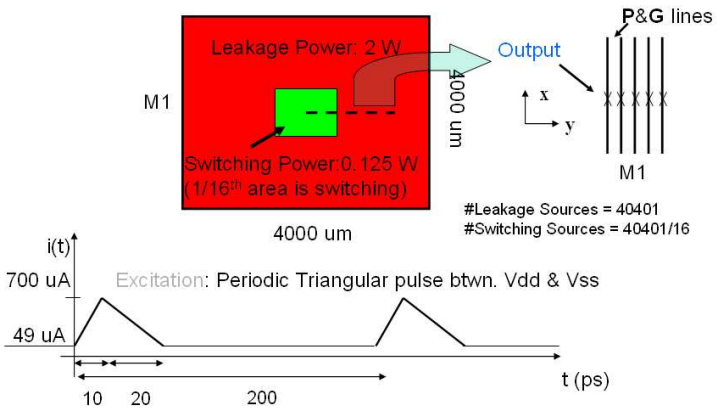
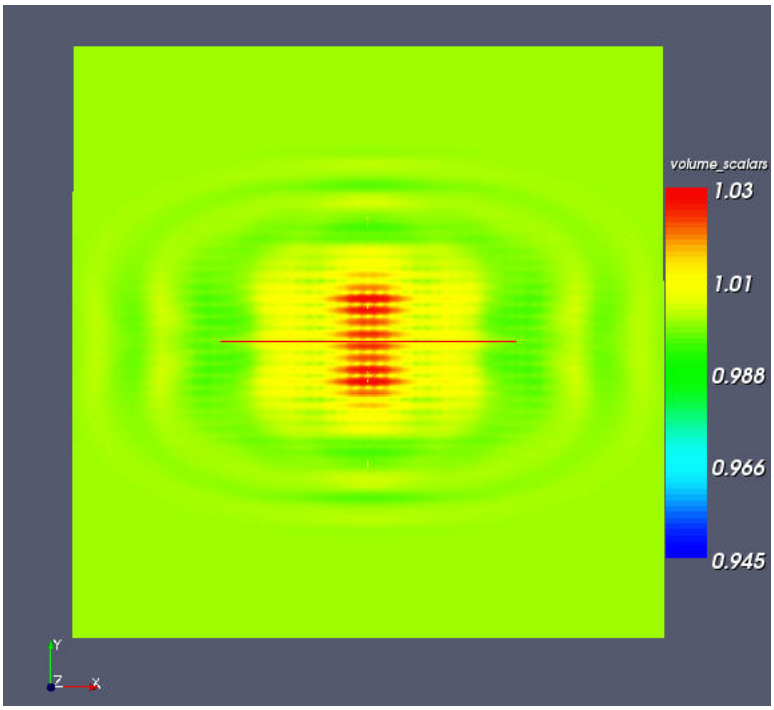



**Figure 31. Switching and leakage current sources and the output node locations.**

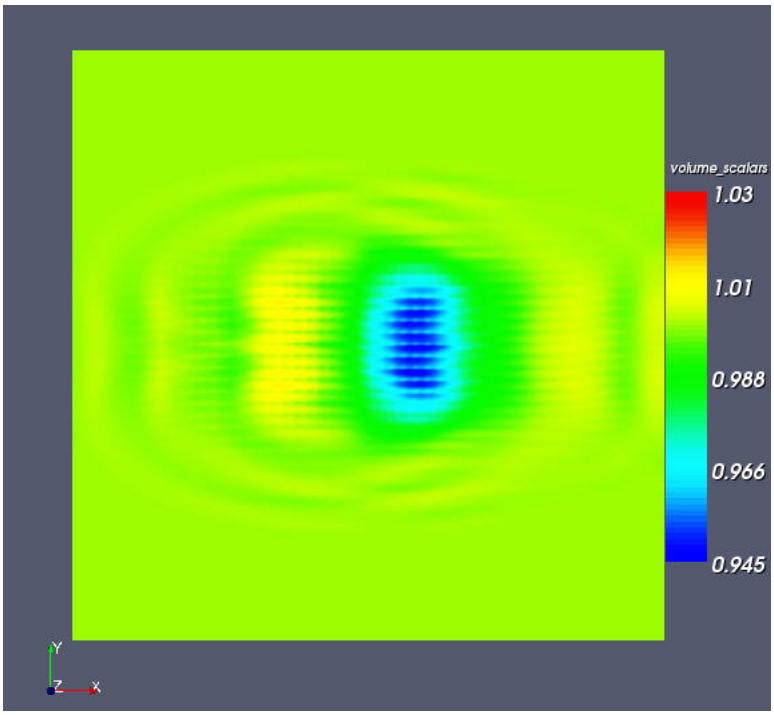
location for the uniform case is 1.00691 V. The difference in their voltages is around 62 mV. This result shows that the nonuniform cross section of lines in the on-chip PDN could affect the PSN. Thus, it might be important to model the various irregularities of the on-chip PDN.

### 5.3.2 Effect of Broken Lines on PSN

To show the effect of the discontinuous lines on the PSN, the following change is made to the test setup. The lines in M1 are made to have the same width and the lines in M2 are made discontinuous by removing the metal in a square region at the center of M2, as shown in Figure 33. The size of this square is varied, and the PSN is computed for the varying sizes. The switching and leakage source descriptions and the output locations are the same as they were in the nonuniform cross-section test case. In Figure 34(a), the transient voltages at the center of M1 are compared with and without the discontinuity in M2. From 34(a), it can be seen that the peak amplitude of the voltage with discontinuous lines (size of discontinuity is 100 $um^2$) is more (about 30 mV) than that of the voltage with continuous lines. In Figure 34(b), the maximum PSNs for varying sizes of the discontinuity are compared. From Figure 34(b), it can be seen that the maximum PSN increases with the increase in the area




(a) Continuous lines.




(b) Discontinuous lines.

**Figure 32. Comparison of the distribution of node voltages in M1 between lines with and without uniform cross section at the end of 35 ps.**

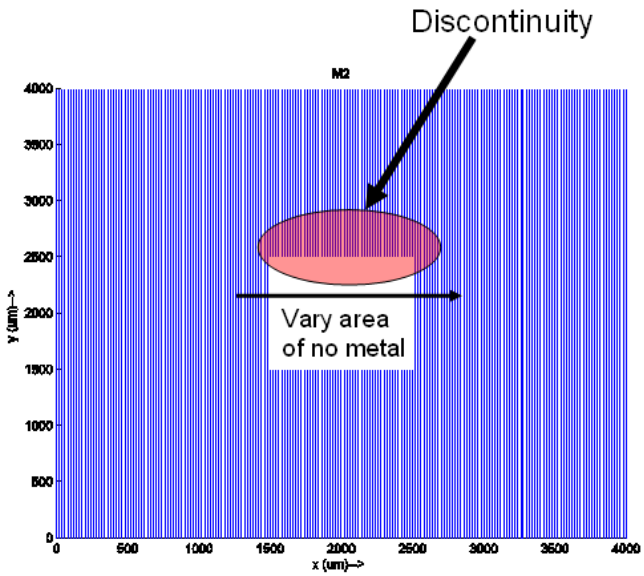
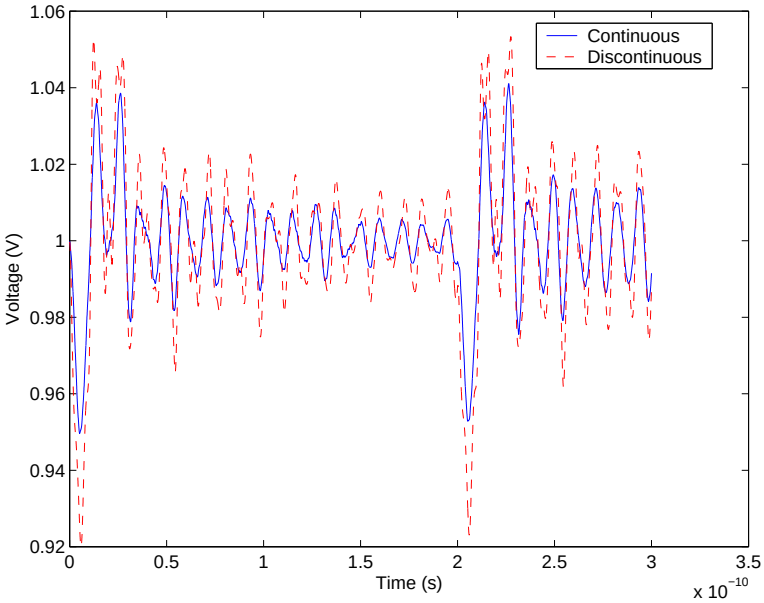



**Figure 33. Geometry of the discontinuous lines in M2. The lines in M2 run parallel to the $y$-axis and have a pitch of 40 um.**

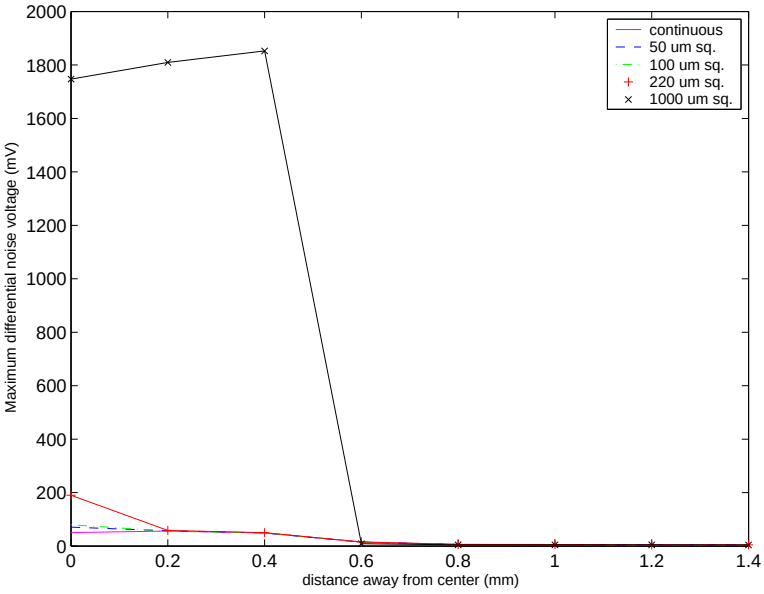
of the discontinuity. The results in Figure 34(a-b) qualitatively verifies the accuracy of the simulator for geometries with discontinuous lines.

## 5.4 Summary

In this chapter, the circuit-FDTD method has been extended to compute the PSN in on-chip PDN geometries where lines can have nonuniform cross sections and be discontinuous. When a line has nonuniform cross section, the parasitic inductance and the parasitic capacitance of the line change. As a result, the Courant time step of the circuit-FDTD method also changes. When the lines become discontinuous, finding the locations of the vias and the crossover capacitors becomes computationally difficult. The accuracy of the simulator has been demonstrated by comparing the PSNs in the on-chip PDN with and without the line irregularities.




(a) Comparison of transient voltages with and without discontinuities: Observation location is at the center of M1; size of the discontinuity in M2 is 100 um$^2$; The voltage with discontinuous lines have more peak-to-peak amplitude than the voltage without discontinuous lines.




(b) Comparison of PSN for increasing discontinuity. The magnitude of the PSN increases with the increase in the size of the discontinuity.

**Figure 34. Effect of the discontinuous lines on the PSN.**

# CHAPTER 6

# ON-CHIP POWER GRID SIMULATION USING LATENCY INSERTION METHOD

## 6.1 Introduction

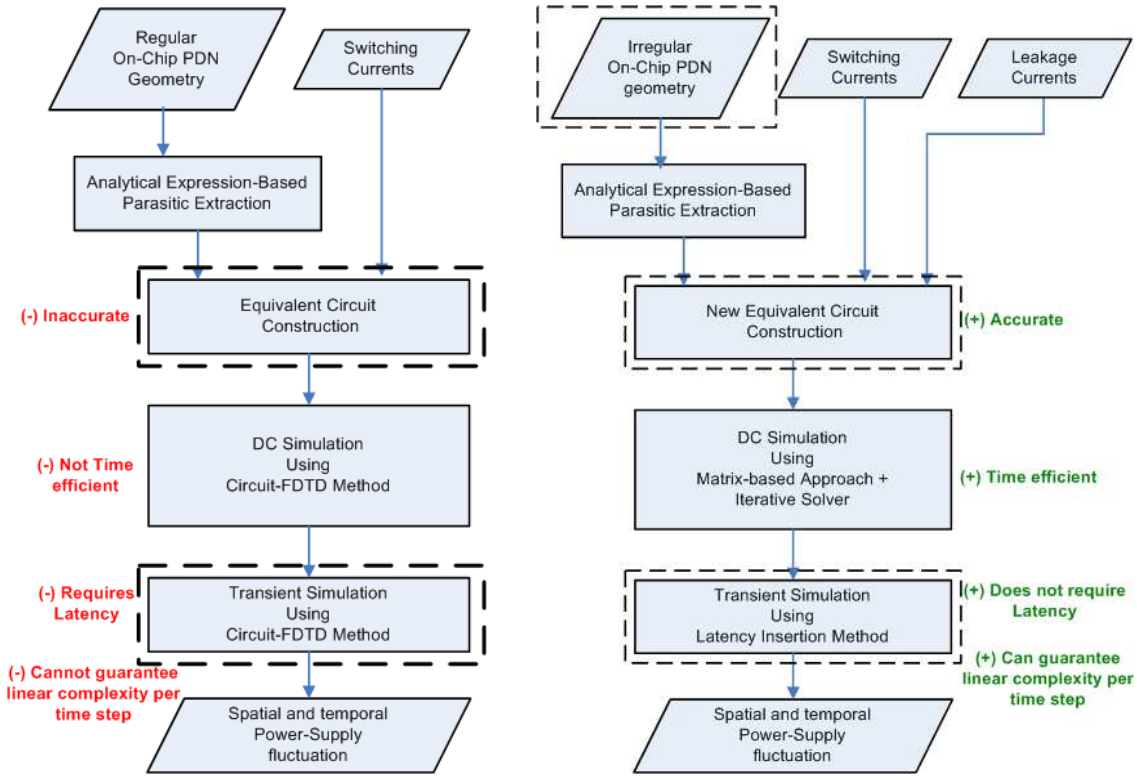
Until now, the circuit-FDTD method has been employed for the power grid transient simulation in [39], [40], [45], in [41], [31], and in Chapters 2, 3, and 5 of this dissertation. In this chapter, it has been demonstrated that to have linear computational complexity per time step of the transient simulation, the circuit-FDTD can no longer be employed; instead, a formulation based on LIM [48] has to be employed. The crossover capacitor formulation proposed in Chapter 3 has been modified to efficiently model any branch capacitance. The focus of this chapter is also described in Figure 35. In Section 6.2, the motivation for using LIM over circuit-FDTD method is described. The contributions and organization of this chapter are described after Section 6.2.5.

## 6.2 Why LIM over Circuit-FDTD Method?

In Sections 6.2.1-6.2.2, the problems with the circuit-FDTD method are described. In Sections 6.2.3-6.2.5, the need for an alternate formulation using LIM has been justified.

### 6.2.1 Capacitance to Ground Problem in On-Chip PDN Equivalent Circuit

The circuit-FDTD method [31], [42] has two problems in the on-chip PDN equivalent circuit that can potentially affect the $O(N_n)$ memory and time complexities per time step of the method. In this chapter, these two problems have been solved while still maintaining the original advantages of the method. In [31], [42], a distributed $\pi-$type RLC equivalent circuit has been employed for modeling power and ground




(a) Prior approach [45], [39], [40], [41], [31].

(b) Proposed approach in this dissertation.

**Figure 35. Comparison of the prior and proposed approach in FDTD-based circuit simulation of PSN in on-chip power grids. The focus of this chapter is the feature in the figure marked within the dashed rectangle.**

lines in the on-chip PDN. Loop-based quantities are used in this equivalent circuit, and the capacitance is dropped to ideal ground. The per-unit-length (p. u. l.) loop resistance, inductance, and capacitance of a line have been extracted assuming that the return currents flow in the lines in the coplanar metal layer and in the lines in the alternate metal layers. Since the lines in the adjacent layers are routed orthogonally, there is no inductive coupling between lines in the adjacent layers; therefore, their effect is ignored in this extraction. There are two problems in this procedure: 1) Though the lines in the adjacent layers do not affect the inductance and resistance extraction, they can affect the capacitance extraction - the lines in the adjacent metal layer can shield the electric flux and hence can block the flux from reaching the lines in the alternate metal layers; therefore, the capacitance is overestimated without considering the effect of the lines in the adjacent layers. 2) Since the capacitance is actually between a line and its return path and the lines comprising the return path are nonideal, the capacitance should not be dropped to ideal ground; instead, it has to be between two nonideal nodes.

### 6.2.2 Computational Inefficiency of Circuit-FDTD Method in Circuits Lacking Latency

These two problems have been addressed by modifying the capacitance extraction procedure and the equivalent circuit. The first problem can be corrected by considering only the capacitances between lines in the coplanar metal layer and between lines in the adjacent metal layers. The second problem can be corrected by having these capacitances between nonideal nodes, i.e., these capacitances are now branch capacitances. Therefore, a new equivalent circuit for the on-chip PDN has been proposed. Two problems were foreseen while simulating the PSN using the approach proposed in [42] (also described in Chapter 3) in the new corrected equivalent circuit. 1) In an interdigitated power-grid (power and ground lines alternate), which is the most common type of power grid, the coplanar line-to-line capacitances and the adjacent-layer

line-to-line capacitances coupled all nodes in a cross-section of the on-chip PDN. As a result, the number of such nodes can be a function of $N_n$. Since in [42], the voltages of nodes that are capacitively coupled are solved simultaneously, a linear system whose size is a function of $N_n$ has to be solved. Since there can be many such sets (the number of sets can also be a function of $N_n$) of capacitively coupled nodes, the node voltages of all nodes are updated by solving many (dependent on $N_n$) linear systems whose sizes are dependent on $N_n$. As a result, the memory and time complexity of updating the voltages of all $N_n$ nodes for a particular time step cannot be strictly guaranteed to be $O(N_n)$. 2) Since the line-to-line capacitances are now floating, there can be some nodes in the new equivalent circuit that are connected to their neighbors only through series resistor-inductor branches and would not have a capacitance to (ideal) ground. For such a node, the Kirchoff's current law (KCL) at the node would not relate the derivative of the node voltage to the currents in the branches connected to the node. Such a relation is essential for updating the node voltages and the branch currents independently and is required by the approaches [39], [31], [42]. In this chapter, these two simulation problems have been overcome while maintaining the $O(N_n)$ memory and time complexity of the simulation for each time step.

### 6.2.3 Need for Inserting Latency Elements

Recently, an explicit simulation approach called the latency insertion method (LIM) [48] has been proposed for large networks in which the nodes need not have capacitance to ground and can be coupled directly to the other nodes either through a resistive or a capacitive branch. In the LIM, a small capacitance to ideal ground is added to a node that did not have a capacitance to ground, and a small series inductance is added to a branch that did not have a series inductance. By adding these extra circuit elements (referred to as fictitious elements in this chapter) the method avoids inverting a large (function of $N_n$) nonbanded system. The formulation [31] and [42] are originally based on LIM, but did not have to use these extra elements.

However, to solve the two simulation problems above, the fictitious elements are added in this chapter. Adding fictitious elements causes two problems, one related to the accuracy and the other related to the time complexity.

### 6.2.4 Need for a Closed-Form Expression for Latency Elements

When fictitious elements are added, accuracy can be comprised. As a result, their values have to be kept small. In the original LIM [48], the values of fictitious elements are computed by repeating the simulation with successively reduced values of these elements until the accuracy is no longer compromised. Such trial-and-error approach to computing these values can get prohibitive in terms of time in circuits with a large $N_n$, especially in on-chip PDN equivalent circuits. To avoid this trial-and-error approach, the values of fictitious elements have to be computed prior to the simulation. However, for generic circuits, such computation is difficult. In this chapter, closed-form expressions have been proposed for computing the values of fictitious elements in the new equivalent circuit of the on-chip PDN. Therefore, fictitious element values are known before the transient simulation, avoiding the cumbersome trial-and-error approach [48] to computing them.

### 6.2.5 Need for Reassessing the Time Complexity with Fictitious Latency Elements

As the values of the fictitious elements are kept small, the maximum time step of the simulation would be affected. Consequently, the number of time steps, $N_t$, can become large. Since $\Delta t$ (and hence $N_t$) is dependent only on the smallest inductance and capacitance values, $\Delta t$ (and hence $N_t$) is independent of $N_n$. Therefore, the time complexity of the transient simulation, $O(N_t N_n)$, is theoretically $O(N_n)$. However, in practice, the time complexity is usually more and is not known quantitatively. This is because a realistic estimate for the time step is not known. However, estimating the time complexity is important in assessing the relative merits/demerits of the approach and in finding methods that improve it. It has been shown in this chapter that the

runtime of the whole transient simulation is approximately proportional to $N_n^{2-2.5}$ for practical problems (defined as $N_n \geq 1$ million).

In this chapter, the solution to the two simulation problems (discussed earlier in this section) that arose in the explicit method [42] because of the changes in the on-chip PDN equivalent circuit have been presented by adding fictitious elements like that in the LIM. The presented solution preserves the original advantages (listed earlier in Section 2.4.1) of the explicit method [42], even for the new on-chip PDN equivalent circuit. Closed-form expressions for computing the values of the fictitious elements have been proposed. Therefore, the fictitious element values can be computed prior to the simulation; hence, the time that would otherwise be incurred in the trial-and-error approach of finding the fictitious element values is avoided. The effect of fictitious elements on the maximum time step has been found. It has been found that the maximum time step is reduced further and this time step can be on the order of femtoseconds. The runtime for the overall simulation has been estimated to be proportional to $N_n^{2-2.5}$ for $N_n$ on the order of millions.

*The contribution of this chapter*[1] *are as follows:*

1. *A new common-mode type equivalent circuit for on-chip power grids.*

2. *An LIM-enabled formulation for the power-grid transient simulation for the new common-mode type equivalent circuit guaranteeing $O(N_n)$ complexity per each time step.*

3. *Using a closed-form-based approach to computing the fictitious element values.*

4. *Getting an estimate for the practical runtime of the proposed LIM-based transient simulation on the proposed equivalent circuit.*

[1] **S. N. Lalgudi**, M. Swaminathan, and Y. Kretchmer, "On-Chip Power Grid Simulation using Latency Insertion Method," Accepted for Future Publication in *IEEE Trans. on Circuits and Systems-I: Fundamental theory and applications*, June 2008.

The rest of this chapter is organized as follows. In Section 6.3, the new equivalent circuit of the on-chip PDN has been described. In Section 6.4, the LIM has been described. In Section 6.5, the proposed LIM-enabled formulation for the power-grid simulation has been described. In Section 6.6, the approximate closed-form expressions for the values of the fictitious elements have been derived. In Section 6.7, the memory and time complexities of the LIM for the on-chip PDN transient simulation have been derived. In Section 6.8, the accuracy of the LIM-enabled transient simulation and the accuracy of the proposed closed-form expressions have been demonstrated. Finally, in Section 6.9, the conclusions have been reported.

## 6.3 Equivalent circuit models of the on-chip PDN and the switching sources

In this section, the new equivalent circuit of the on-chip PDN is described. A simplified 3-D view of an on-chip PDN with three metal layers shown in Figure 36 has been considered for the simulation. The equivalent circuit of the on-chip PDN shown in Figure 36 has been shown in Figure 37. The changes to the equivalent circuit already described is only discussed here.

From Section 6.2, two kinds of capacitances have to be considered instead of the capacitance to ideal ground. These capacitances are the crossover capacitance and the coplanar line-to-line capacitances. Crossover capacitance modeling is same as the one described in Chapter 3.

Among all coplanar layer line-to-line capacitances, only the capacitances between adjacent lines are more important than the rest. Therefore, coplanar capacitances between nonadjacent lines are ignored. The line-to-line capacitance between a power line and its nearest ground line in the coplanar layer is directly proportional to their thicknesses and is inversely proportional to the distance between them. Since the distance between a power line and its nearest ground line in a metal layer (typical

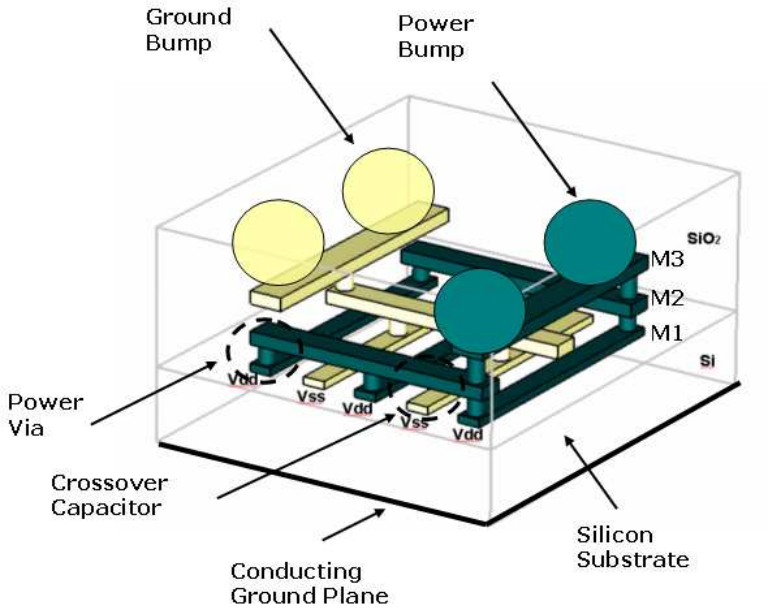



**Figure 36. Simplified 3-D view of an on-chip power distribution network with 3 metal layers; M1 is the metal layer closest to the silicon substrate; M3 is the metal layer farthest from the substrate; and M2 is the metal layer between M1 and M3.**

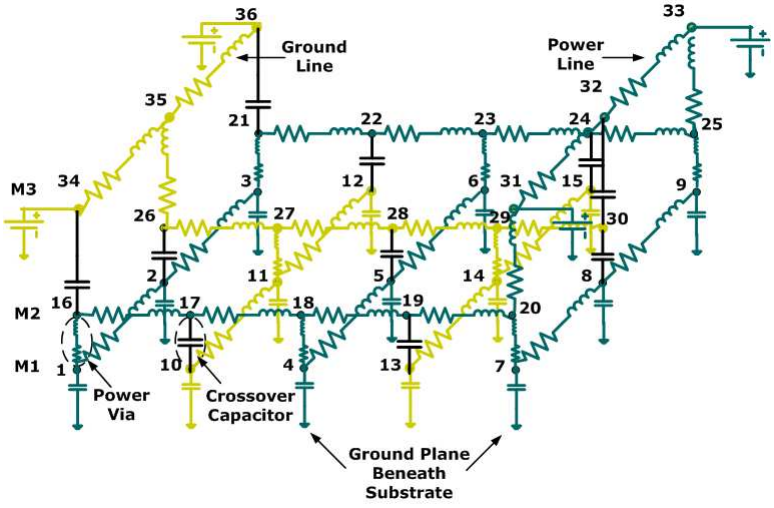



**Figure 37. The equivalent circuit of the on-chip PDN shown in Figure 36.**

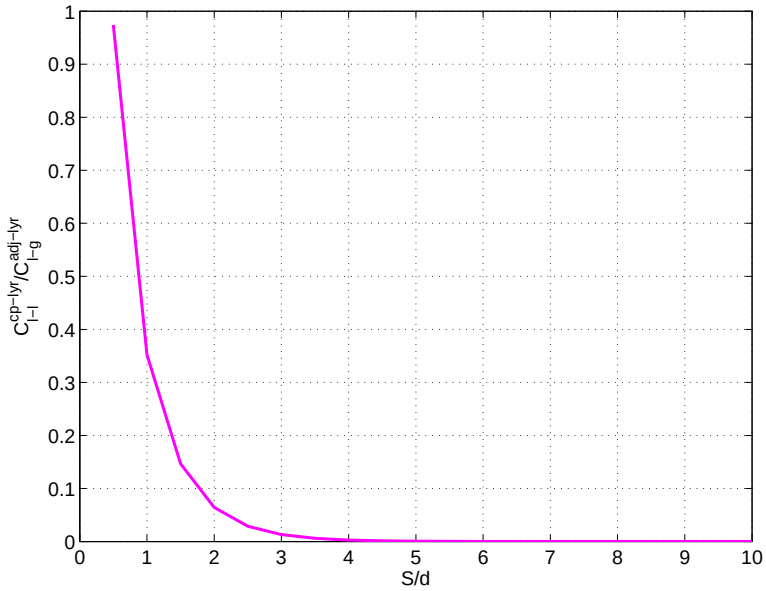



**Figure 38. Comparison of coplanar line-to-line capacitance and adjacent layer line-to-ground capacitance. d is the distance between metal layers; $C_{\text{l-l}}^{\text{adjlyr}}$ is the capacitance per-unit-length between two lines in the same layer separated by distance $S$; $C_{\text{l-l}}^{\text{cplyr}}$ is the capacitance per-unit-length between a line and the adjacent metal layers (modeled as solid planes) at a distance $d$.**

value may lie between 10 and 100 um) is usually much greater than the distance between adjacent metal layers (typical values may lie between 1 and 5 um), the coplanar line-to-line capacitance is usually very small compared to the crossover capacitance and therefore is not modeled. In Figure 38, the ratio between the line-to-line capacitance in the coplanar layer and the line-to-ground capacitance between a line and the adjacent metal layers (which almost acts as a solid plane) has been plotted for different $S/d$. For $S/d \geq 3$, the $C_{\text{l-l}}^{\text{adjlyr}}$ / $C_{\text{l-l}}^{\text{cplyr}} \leq 0.01$.

Besides the line-to-line capacitances described above, there is also a line-to-ground capacitance (see Figure 37) between lines in M1 and the conducting ground plane (assumed to be ideal) beneath the silicon substrate. Apart from the line capacitances described above, there are built-in and added on-chip decoupling capacitances [7]. These capacitances are not modeled in this paper. Moreover, the conductivity of the substrate could affect the DC voltages of the ground nodes [73]. However, in this work, this effect is not modeled.

Besides the coupling capacitances, there are also coupling inductances between lines. Since the coupling inductance between two lines is inversely proportional to the distance between them, the error introduced by ignoring coupling between distant lines would not be much. Ignoring the far-away coupling and considering only the nearby coupling are fraught with stability issues [74] and should be dealt with caution (see [75] and the references therein). In this work, though an uncoupled inductance model is proposed, the coupling between neighboring lines is partially addressed. This is because the inductance is a loop inductance (which is function of both the self and the mutual partial inductances) assuming nearby return paths (see [31] for more details regarding the inductance extraction), so the effect of coupling between nearby lines are already taken into account. Since only the self loop inductance is used in the model, the simulation does not suffer from the stability issues.

The description of the active circuits in the PSN simulation remains the same as that described in Chapter 4.

## 6.4 Latency Insertion Method (LIM)

In this section, the LIM has been described. LIM is same as the circuit-FDTD method in all respects except for the artificial latency elements that might be inserted in the former. The rules governing this insertion are as follows. A branch in a circuit is defined as a connection between two nodes excluding the ground reference node. To enable LIM in a circuit, 1) each branch in the circuit should have a nonzero inductance; otherwise, a small inductance is inserted into the branch to generate latency. 2) each node in the circuit should have a capacitance to ground; otherwise, a small shunt capacitance is added to generate latency at that node.

The LIM is developed to simulate the high-frequency response of a large network in the time domain. In this method, a finite-difference formulation is used to update branch currents and node voltages in a leapfrog manner similar to the Yee algorithm

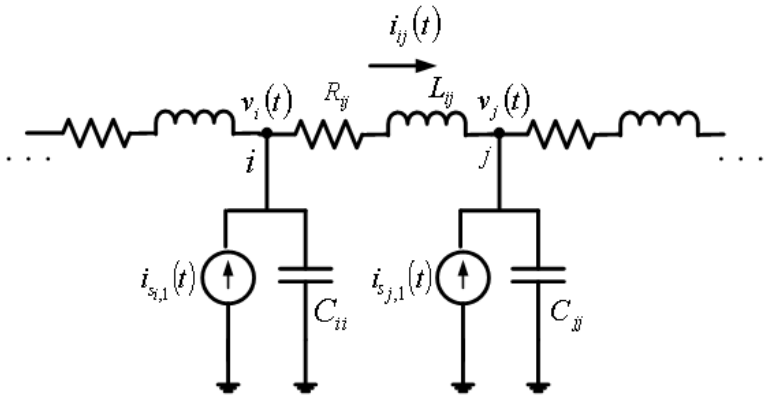



**Figure 39. Typical equivalent circuit to enable LIM.**

used in the finite-difference time-domain (FDTD) method [46]. As a result, the LIM has linear computational complexity. The LIM is readily enabled in networks with latency. A network has latency if each node in it has a shunt capacitance to ground and each branch in it has a series inductance. Such networks are observed in distributed RLC-based transmission line circuits. If latency is missing in some parts of the network, then latency is inserted to enable the LIM. Like the FDTD method, the LIM has an upper bound on the time step, dictated by stability requirements of the update algorithm. In the rest of the section, the formulation, accuracy, computational complexity, and stability of the LIM have been described for a network containing linear sources.

In Figure 39, a sample circuit is shown for which the LIM is enabled. This type of circuit is also common in the proposed equivalent circuit of on-chip power grids. The symbols in Figure 39 mean the following: $i$ and $j$ refer to the nodes; the subscript $ij$ refers to a branch between nodes $i$ and $j$; $R_{ij}$ and $L_{ij}$ are the series resistance and inductance of the branch between nodes $i$ and $j$, respectively; $C_{ii}$ refers to the shunt

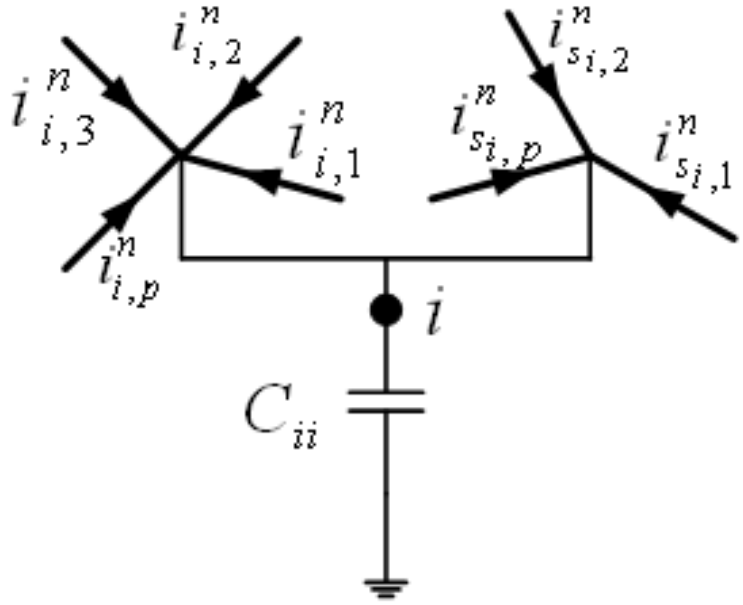



**Figure 40. Conceptual equivalent circuit at node $i$.**

capacitance from node $i$ to ideal ground; $v_i(t)$ refers to the voltage at node $i$ and time $t$; $i_{ij}(t)$ refers to the current in the branch between nodes $i$ and $j$; $i_{s_{i,p}}(t)$ refers to the current as a result of the $p$th current source connected to node $i$. Leapfrog scheme is a second-order integration method to solve differential equations. This scheme relies on staggering the voltages and the currents by half a space step and half a time step. By defining currents for all branches and voltages for all nodes, the spatial staggering needed for the leapfrog scheme is accomplished. By defining branch currents (also source currents) at time instants $n\Delta t$ and node voltages at $\left(n+\frac{1}{2}\right)\Delta t$, where $n = 0, 1, \ldots,\ N_t - 1$, the temporal staggering needed for the leapfrog scheme is also met. The leapfrog scheme is referred to as a semi-implicit scheme in [49]. LIM can also be formulated using first-order schemes like fully explicit and fully implicit schemes [49].

In the LIM, the transient simulation is accomplished by updating the node voltages and the branch currents at each time step. These expressions are derived for the

circuit shown in Figure 39, starting with the update expressions for the node voltages. The conceptual equivalent circuit at node $i$ in Figure 39 looks like as shown in Figure 40. In Figure 40, $i^n_{i,p}$ refers to the $p$th branch current entering node $i$ at time $t = n\Delta t$; and $i^n_{s_{i,p}}$ is the $p$th source current entering node $i$ at time $t = n\Delta t$. From the Kirchoff's current law (KCL) at node $i$, the equation

$$C_{ii}\frac{dv_i(t)}{dt} = \sum_{p\ =1}^{N_b^i} i_{i,p}(t) + \sum_{p\ =1}^{N_s^i} i_{s_{i,p}}(t) \tag{24}$$

can be obtained, where $N_b^i$ is the number of branches incident on node $i$, and $N_s^i$ is the number of sources incident on node $i$. When the derivative in (24) is discretized at time $t = n\Delta t$, the voltage at time $t = (n + \frac{1}{2})\Delta t$, $v_i^{n+\frac{1}{2}}$, is given by

$$v_i^{n+\frac{1}{2}} = v_i^{n-\frac{1}{2}} + \frac{\Delta t}{C_{ii}}\Big(\sum_{p\ =1}^{N_b^i} i^n_{i,p} + \sum_{p\ =1}^{N_s^i} i^n_{s_{i,p}}\Big). \tag{25}$$

In (25), $v_i^{n+\frac{1}{2}}$ is expressed in terms of only the quantities known at $t = n\Delta t$. Consequently, (25) is an explicit expression for updating the voltage for any node $i$ whose conceptual equivalent circuit is as shown in Figure 40.

The update expression for the branch currents can be derived following a procedure similar to that of the node voltages. The equivalent circuit of a branch is shown in Figure 41. When the Kirchoff's voltage law (KVL) is applied along this branch, the equation

$$v_i(t) = R_{ij}i_{ij}(t) + L_{ij}\frac{di_{ij}(t)}{dt} + v_j(t) \tag{26}$$

can be obtained. When the derivative $\frac{di_{ij}(t)}{dt}$ in (26) is discretized at time $t = (n+\frac{1}{2})\Delta t$, an explicit update expression for $i_{ij}^{n+1}$ can be obtained as

$$i_{ij}^{n+1} = \frac{1 - \frac{R_{ij\Delta t}}{2L_{ij}}}{1 + \frac{R_{ij\Delta t}}{2L_{ij}}} i_{ij}^n + \frac{\frac{\Delta t}{L_{ij}}}{1 + \frac{R_{ij\Delta t}}{2L_{ij}}}\Big(v_i^{n+\frac{1}{2}} - v_j^{n+\frac{1}{2}}\Big). \tag{27}$$

At each time step in the transient simulation, first, all node voltages are updated through (25), and next, all branch currents are updated through (27). The accuracy

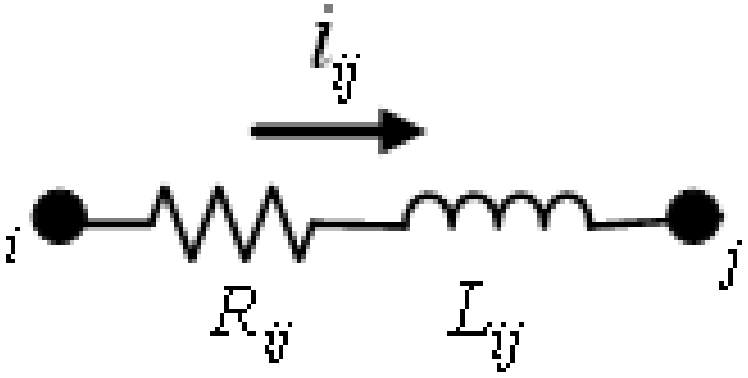



**Figure 41. The equivalent circuit of a branch between node $i$ and $j$.**

of the transient solution scales as $O((\Delta t)^2)$. The memory complexity of the LIM is $O(N_n + N_b)$, and its time complexity is $O\left(N_t\left(N_n + N_b\right)\right)$, where $N_b$ is the total number of branches in the network, hence yielding an optimally efficient algorithm. The time step, $\Delta t$, in the LIM has an upper bound. This restriction on the time step follows from the need to keep the simulation numerically stable. In Chapter 8, analytical stability condition is derived for inhomogeneous RLC and GLC circuits. These conditions in turn leade to an upper bound on the time step.

## 6.5 On-Chip Power Grid Transient Simulation using LIM

To simulate the temporal fluctuation in the power supply as a result of switching sources, a transient simulation is preferred. In this section, a new LIM-enabled formulation for this transient simulation is described. The principal advantage of this new formulation is that it guarantees $O(N_n)$ computational complexity per time step of the transient simulation. This advantage is realized through the artificial insertion of latency in the circuit, which is the difference from the formulations in [39], [31] and [42].

The on-chip power grid shown in Figure 36, which has an equivalent circuit as shown in Figure 37 has been used a reference for describing the formulation. First, it can be noticed that many parts of the equivalent circuit shown in Figure 37 has a

form similar to the one in shown in Figure 39. The current source in Figure 39 can be used to represent the contribution of both the switching and leakage currents. The power and ground voltage supplies at C4 locations are taken into account by enforcing these supply voltages for the node voltages corresponding to the C4 locations. Owing to this similarity, the update expressions developed in the LIM can be readily used in most cases. In fact, in Figure 37, the nodes in M1 at the end points of vias, namely the nodes 1, 3, 4, 6, 7, 9, 11, and 14, have the same conceptual equivalent circuit as in Figure 40; hence, their voltages can be updated using (25).

However, for the rest of the nodes in Figure 37, the node voltages cannot be updated using (25), as the latency is missing in all these nodes. Latency is missing because these nodes either do not have a shunt capacitance to ground or have branch capacitors connected to them. To see why latency is important for updating the node voltage, consider node 18 in M2. This node does not have a shunt capacitance to ground, for reasons described earlier in this paper. The conceptual equivalent circuit at this node is same as in Figure 40 except for $C_{ii}$. If $C_{ii}$ is missing in Figure 40, the KCL equation in (24) would not involve $\frac{dv_i(t)}{dt}$, and the explicit expression in (25) cannot be obtained in the first place.

In [31], [42] and in Chapters 2-5, a shunt capacitance to ground is assumed to be present in all nodes in the on-chip PDN. However, as described earlier in this paper, this assumption may not true. Therefore, the formulation presented in these approaches cannot be applied to the new equivalent circuits shown in Figure 37.

To enable LIM, latency is inserted. For nodes with a missing capacitance to ground (nodes in M2 and above), a small fictitious shunt capacitance to ground is added. With this addition, the equivalent circuit shown in Figure 37 is modified as shown in Figure 42. In the modified equivalent circuit, even the branch capacitors are changed (more about this later).

With the modified equivalent circuit (see Figure 42), the nodes in M2 and above

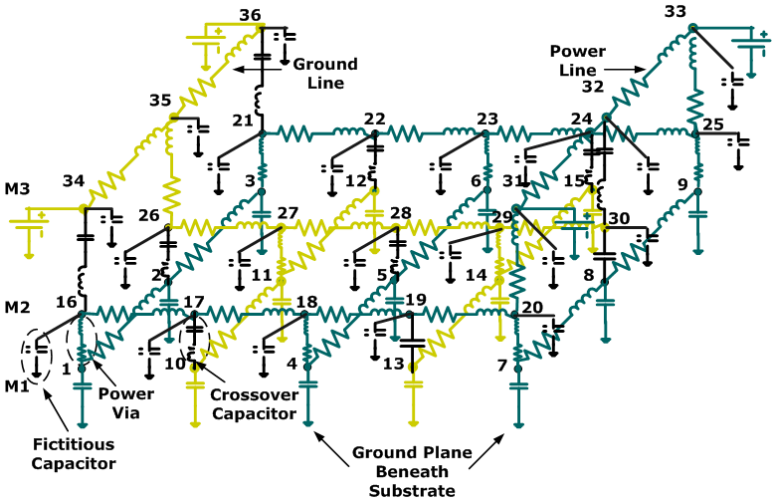



**Figure 42. The equivalent circuit of the on-chip PDN shown in Figure 36 with fictitious elements; fictitious capacitance to ground is added to nodes in M2 and M3 and a fictitious series inductance is added each crossover capacitor.**

that are not connected to any branch capacitors, namely the nodes 18, 20, 23, 25, 27, and 29 in M2, and the nodes 29, 33 and 35 in M3, have the same conceptual equivalent circuit shown in Figure 40. Therefore, their voltages can be updated using (25). Adding a fictitious shunt capacitance to ground affects accuracy, so the values of the fictitious elements have to be small. The choice for the values of fictitious capacitance to ground is described in Section 6.6.

When crossover capacitors are not modeled, there will be no branch capacitors in the equivalent circuit shown in Figure 42. In such a scenario, the rest of the nodes, which are also the end points of the (missing) crossover capacitors, too have the same conceptual equivalent circuit as shown in Figure 40. Therefore, even their voltages can be updated using (25).

When crossover capacitors are modeled, the rest of the nodes are connected either to one or to two branch capacitors. In such a scenario, the conceptual equivalent circuits at these nodes are not the same as the one in Figure 40. Therefore, their voltages cannot be updated using (25), necessitating the development of a new update expression for these nodes.

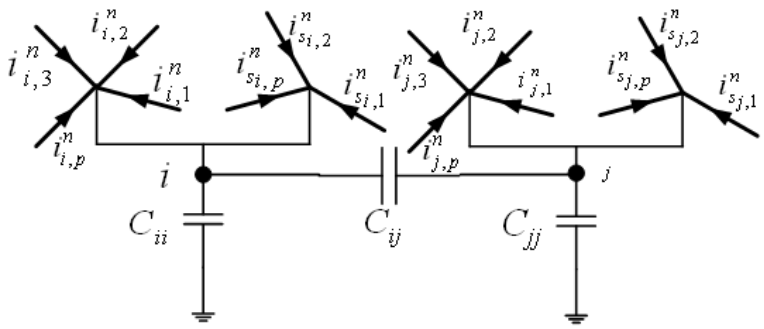



**Figure 43. The conceptual equivalent circuit at the two end nodes $i$ and $j$ of the crossover capacitor $C_{ij}$.**

Towards this end, a conceptual equivalent circuit at the end points of a branch capacitor is considered. In Figure 43, a typical case is shown. There are two ways to derive the update expressions for the voltages of nodes $i$ and $j$ in Figure 43. These ways differ in the decision to insert a latency (by way of a small series inductance) in the branch capacitor. When the latency is not introduced in the branch capacitors, the linear complexity per time step of the approaches [31], [42] may not be guaranteed (shown below). However, when the latency is introduced, this complexity can be guaranteed. Both the ways are described next.

In the first way, latency is not introduced in branch capacitors. This is the way adopted in [31], [42]. The update expressions for the voltages for nodes $i$ and $j$ in Figure 43 are derived as follows. The process starts with the KCL at node $i$. The KCL at node $i$ would involve terms $\frac{dv_i(t)}{dt}$ and $\frac{dv_j(t)}{dt}$. When these derivatives are discretized, $v_i^{n+\frac{1}{2}}$ is given by

$$\Big(C_{ii}+C_{ij}\Big)v_i^{n+\frac{1}{2}} - C_{ij}v_j^{n+\frac{1}{2}} = \Big(C_{ii}+C_{ij}\Big)v_i^{n-\frac{1}{2}} - \\ -C_{ij}v_j^{n-\frac{1}{2}} + \Delta t\Big(\sum_{p\ =1}^{N_b^i} i_{i,p}^n + \sum_{p\ =1}^{N_s^i} i_{s_{i,p}}^n\Big). \qquad (28)$$

In (28), $v_i^{n+\frac{1}{2}}$ is related to the unknown quantity $v_j^{n+\frac{1}{2}}$. Hence, $v_i^{n+\frac{1}{2}}$ cannot be obtained using (25). In fact, both $v_i^{n+\frac{1}{2}}$ and $v_j^{n+\frac{1}{2}}$ have to be solved together. The extra equation needed to find this solution is obtained from the KCL equation at

node $j$ and is given by

$$-C_{ij}v_i^{n+\frac{1}{2}} + \Big(C_{ii} + C_{ij}\Big)v_j^{n+\frac{1}{2}} = -C_{ij}v_i^{n-\frac{1}{2}} +$$

$$+\Big(C_{ii} + C_{ij}\Big)v_j^{n-\frac{1}{2}} + \Delta t\Big(\sum_{p\ =1}^{N_b^j} i_{j,p}^n + \sum_{p\ =1}^{N_s^j} i_{s_{j,p}}^n\Big). \tag{29}$$

Equations (28) and (29) are solved together to find $v_i^{n+\frac{1}{2}}$ and $v_j^{n+\frac{1}{2}}$. When node $i$ ($j$) in Figure 43 is also connected capacitively to some other node in the PDN, the size of the system to be solved increases, increasing the computational complexity of updating these node voltages. However, in [42] and in Chapter 3, it has been shown that the linear computational complexity per time step is preserved when only crossover capacitances are considered.

However, when decoupling capacitors are present, the linear computational complexity may be compromised using this way. When the on-chip decoupling capacitors are spread uniformly across the chip, the number of nodes that are capacitively coupled could be proportional to $N_n$. When the number of capacitively coupled nodes is proportional to $N_n$, a large sparse system whose size is proportional to $N_n$ needs to be solved. The complexity of the solution to such a system cannot be guaranteed to be linear, necessitating efforts to avoid this complexity problem.

In the second way, latency is inserted in all branch capacitors (whether crossover or decoupling capacitors) by adding a small series inductance to them. The choice for the series fictitious inductance in the crossover capacitor is described in Section 6.6. Consequently, the crossover capacitor $C_{ij}$ is represented by a series resistor-inductor-capacitor model as shown in Figure 44. Further, the current through the crossover capacitor as well as the voltage of its internal node (node $k$ in Figure 44) are maintained. In Figure 42, the on-chip PDN equivalent circuit with the modified crossover capacitor model has been shown. With the new crossover capacitor model, the conceptual equivalent circuit shown in Figure 43 becomes like the one in Figure 45. Since the current $i_{ij}$ (see Figure 45) through the crossover capacitor is maintained,

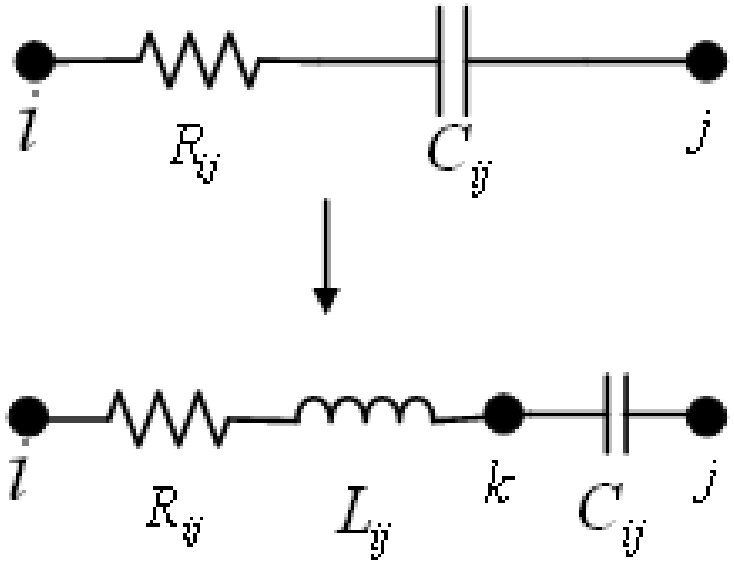



**Figure 44. The new equivalent circuit of a floating capacitor $C_{ij}$.**

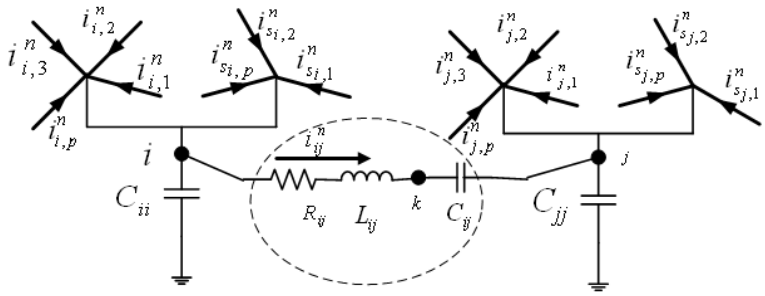



**Figure 45. The conceptual equivalent circuit at the two end nodes $i$ and $j$ with the new model for the crossover capacitor $C_{ij}$.**

the conceptual equivalent circuit at node $i$ is same as the one shown in Figure 40. Therefore, $v_i^{n+\frac{1}{2}}$ of node $i$ in Figure 45 can be obtained using the expression in (25). As a result, $v_i^{n+\frac{1}{2}}$ is solved independently of $v_j^{n+\frac{1}{2}}$.

However, a similar procedure cannot be employed for obtaining $v_j^{n+\frac{1}{2}}$. Since node $j$ is capacitively connected to node $k$, their voltages have to be solved together using the procedure described in the first way. However, unlike in the first way, the linear computational complexity per time step can be guaranteed: The size of the system to be solved for updating a node voltage is equal to the number of branch (floating) capacitors connected to the node. Since only the capacitive coupling between a node and its neighbors is modeled in the equivalent circuit shown in Figure 37, the maximum size of the system to be solved is equal to the number of neighboring nodes of a node. This number is independent of $N_n$ (for equivalent circuits shown in Figure 37), even if on-chip decoupling capacitors were to be included.

Unlike the procedure for updating the node voltages, the procedure for updating the branch current is relatively simple. As all branches, including the branch capacitors, have the same conceptual equivalent circuit as shown in Figure 41, the branch currents are updated using (27).

The time step of the transient simulation has an upper bound. In Chapter 8, this upper bound is derived for inhomogeneous RLC and GLC circuits. A more strict upper bound for the time step than (68) (in Chapter 8) is used:

$$\Delta t \leq \min_{i,p} \left( \frac{\sqrt{L_{i,p} C_{ii}}}{\sqrt{N_b^i}} \right), \tag{30}$$

where ${N_b}^i$ is the total number of branches connected to node $i$. The time step in (30) is observed to be working even if there are coupling capacitances in the equivalent circuit.

## 6.6 Closed-Form Expressions for Fictitious Latency Elements

In this section, approximate closed-form expressions for computing the fictitious latency elements have been derived.

### 6.6.1 Fictitious Series Inductance

The fictitious series inductance added to a floating capacitor (see Figure 44) makes this capacitance a series inductor-capacitor resonance circuit. The objective is to choose the fictitious inductance $L_{ij}^f > 0$ such that it does not affect the accuracy of the results (obtained without it) much. At low frequencies, $L_{ij}^f$ acts as a short circuit, and its effect is not felt. However, at high frequencies, its impedance is high; hence, it can affect the results. The objective stated above will be met if $L_{ij}^f$ is chosen such that even at the maximum frequency of interest ($f_{\max}$), the inductance has a significantly smaller impedance compared to the impedance of the rest of the capacitor circuit; i.e.,

$$\omega_{\max} L_{ij}^f << \sqrt{R_{ij}^2 + \frac{1}{(\omega_{\max} C_{ij})^2}},$$

or

$$\omega_{\max} L_{ij}^f = k_L \sqrt{R_{ij}^2 + \frac{1}{(\omega_{\max} C_{ij})^2}}, \tag{31}$$

where $k_L \ll 1$, and $\omega_{\max} = 2\pi f_{\max}$. Then, the fictitious series inductance can be computed from

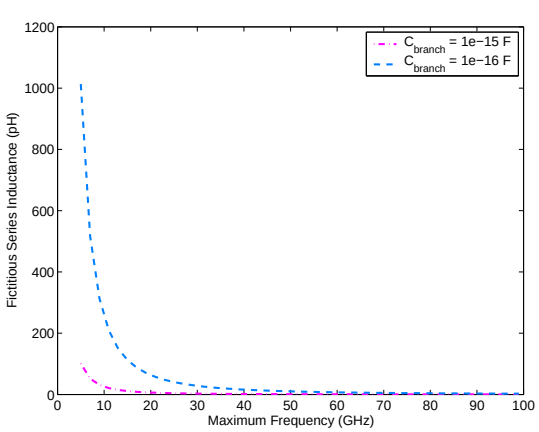
$$L_{ij}^f = \frac{k_L}{\omega_{\max}} \sqrt{R_{ij}^2 + \frac{1}{(\omega_{\max} C_{ij})^2}}. \tag{32}$$

From (32), it can be observed that the fictitious inductance decreases with the increase in $f_{\max}$, with the increase in $C_{ij}$, and with the decrease in $k_L$. The smaller the fictitious series inductance, the more accurate is the result. The factor $k_L$ can be employed to control the accuracy. Since time step $\Delta t$ depends on the fictitious series inductance (see (30)), it would be useful to compute this inductance for some practical values of $f_{\max}$ and $C_{ij}$. If $f_{\max} = \frac{1}{t_r}$, where $t_r$ is the rise time of the (triangular) switching current source, then 5 GHz $\leq f_{\max} \leq$ 100 GHz for 200 ps $\geq t_r \geq$ 10 ps.

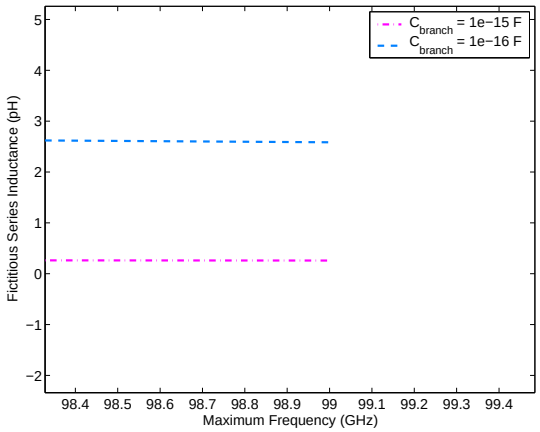
The crossover capacitance is usually less than 10 fF (The crossover capacitance, $C_{cr}$, can be approximately computed by $C_{cr} = \varepsilon_{SiO_2} \frac{w_1 w_2}{d}$, where $\varepsilon_{SiO_2}$ is the permittivity of silicon dioxide; $w_1$ and $w_2$ are the widths of lines in adjacent metal layers, and $d$ is the distance between the adjacent metal layers; if $w_1 = 8$ um, $w_2 = 10$ um, and $d = 2$ um, then $C_{cr} = 1.7$ fF.) The thin-oxide on-chip decoupling capacitor for 1 sq-um area with a gate length of 90 nm and with an oxide thickness of 1.3 nm is approximately 26 fF. Therefore, the fictitious series inductance, $L_{ij}^f$, is computed using (32) for 5 GHz $\leq f_{\max} \leq$ 100 GHz, 0.1 fF $\leq C_{ij} \leq$ 100 fF, $k_L = 10^{-3}$, and $R_{ij} = 0$. In Figure 46, the plots for $C_{ij} = 0.1$ fF, 1 fF, 10 fF, and 100 fF are shown. From Figure 46, it can be observed that the fictitious series inductance can be low or high depending on the values of $C_{ij}$ and $f_{\max}$. This inductance has been found to be as low as 2.5 fH for $C_{ij} = 100$ fF and $f_{\max} = 100$ GHz, and as high as 1000 pH for $C_{ij} = 0.1$ fF and $f_{\max} = 5$ GHz.

### 6.6.2 Fictitious Capacitance to Ground

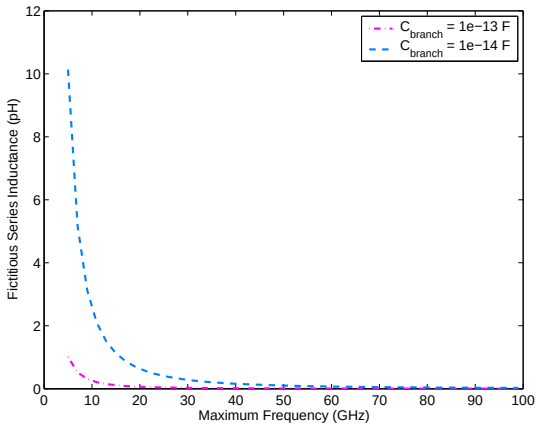
The procedure for computing the fictitious capacitance to ground (in a node that did not have it before) is similar to the one followed for computing the fictitious series inductance. Since all nodes in power (ground) rails of the on-chip PDN will have a DC path to the power (ground) supply, the equivalent circuit from node $i$ (with no capacitance to ground) to the power (ground) supply can be represented as in Figure 47(a). In Figure 47(a), $V_s$ is the power/ground supply voltage source, $R_{i-s}$ and $L_{i-s}$ are the resistance and inductance, respectively, between node $i$ and the voltage source, and $Z_{i-s}(\omega)|_{w/o\,C}$ is the impedance observed from node $i$ and looking into the DC voltage source in the absence of capacitance to ground from node $i$. When a fictitious capacitance to ground, $C_{ii}^f$, is added, the equivalent circuit shown in Figure 47(a) can be transformed to the one in Figure 47(b). The objective is to choose $C_{ii}^f > 0$ in such a way that it does not affect the accuracy of the results (obtained without it) much. At low frequencies, $C_{ii}^f$ acts as an open circuit; therefore, its effect is not felt.



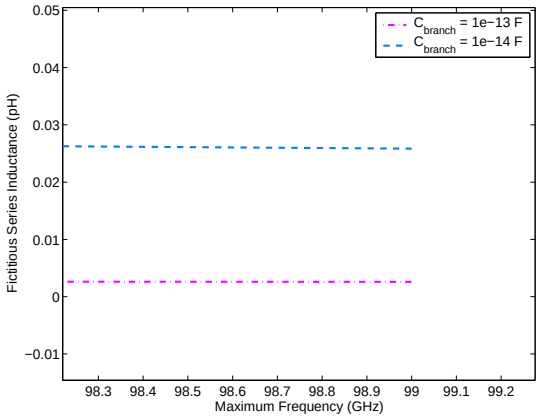

(a) Branch capacitance of 0.1 and 1 fF




(b) Branch capacitance of 0.1 and 1 fF

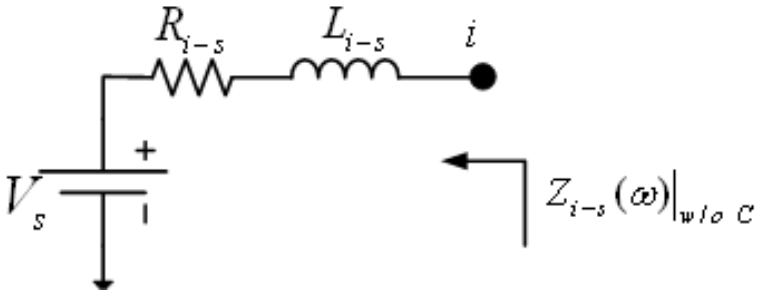



(c) Branch capacitance of 10 and 100 fF

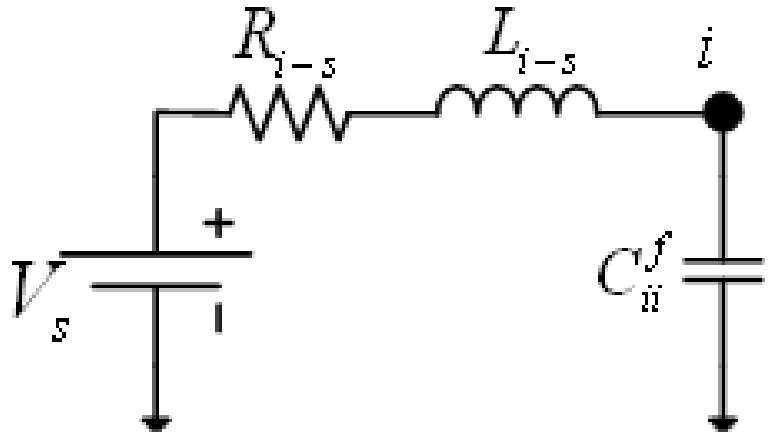



(d) Branch capacitance of 10 and 100 fF

**Figure 46. Variation of fictitious inductance with maximum frequency of the excitation and with floating capacitance.**




(a) Without fictitious capacitance to ground




(b) With fictitious capacitance to ground

**Figure 47. The equivalent circuit as seen from node $i$ in an on-chip PDN to the power/ground supply terminal. $V_s$ is the power/ground supply; $R_{i-s}$ and $L_{i-s}$ are the net resistance and inductance, respectively, between node $i$ and supply voltage; and $C_{ii}^f$ is the fictitious capacitance to ground from node $i$.**

However, at high frequencies, $C_{ii}^f$ may have a smaller impedance than $Z_{i-s}(\omega)|_{w/o\,C}$ and its effect might be felt. The impedance $Z_{i-s}(\omega)|_{w/o\,C}$ is given by

$$Z_{i-s}(\omega)|_{w/o\,C} = R_{i-s} + I\omega L_{i-s} \tag{33}$$

To not affect $Z_{i-s}(\omega)|_{w/o\,C}$, $C_{ii}^f$ can be chosen such that it presents a much higher impedance than $\left| Z_{i-s}(\omega)|_{w/oC} \right|$ at $f_{\max}$; i.e.,

$$\frac{1}{\omega_{\max} C_{ii}^f} >> \sqrt{R_{i-s}^2 + (\omega_{\max} L_{i-s})^2},$$

or

$$\frac{1}{\omega_{\max} C_{ii}^f} = \frac{1}{k_C}\sqrt{R_{i-s}^2 + (\omega_{\max} L_{i-s})^2}, \tag{34}$$

where $k_C \ll 1$. The capacitance to ground $C_{ii}^f$ is then computed from

$$C_{ii}^f = \frac{k_C}{\omega_{\max}\sqrt{R_{i-s}^2 + (\omega_{\max} L_{i-s})^2}} \tag{35}$$

for all nodes $i$ that do not have a capacitance to ground. Since $R_{i-s}$ and $L_{i-s}$ vary depending on node $i$ and are difficult to compute at each node $i$, maximum of their values can be used. Accordingly, the fictitious capacitance to ground, $C_{ii}^f$, can be computed for any node $i$ by

$$C_{ii}^f = \frac{k_C}{\omega_{\max}\sqrt{\left(\max_i \{R_{i-s}\}\right)^2 + \left(\omega_{\max} \max_i \{L_{i-s}\}\right)^2}}. \tag{36}$$

If there are more than one power (ground) supply, then only the nearest power (ground) supply is considered. Approximate estimates for terms $\max_i \{R_{i-s}\}$ and $\max_i \{L_{i-s}\}$ in (36) has been derived below.

The distance between a node and its nearest power-(ground-) supply bump is usually bounded by the pitch of the bumps. This has been illustrated in Figure 48. In Figure 48, the typical bump arrangement has been shown. In Figure 48, $\mathrm{B}_i$ refers to the $i$th power (ground) bump, and $S_{\mathrm{bx}}$ and $S_{\mathrm{by}}$ refer to the spacing between adjacent power (ground) bumps in the $x$- and the $y$- directions, respectively. Then, the

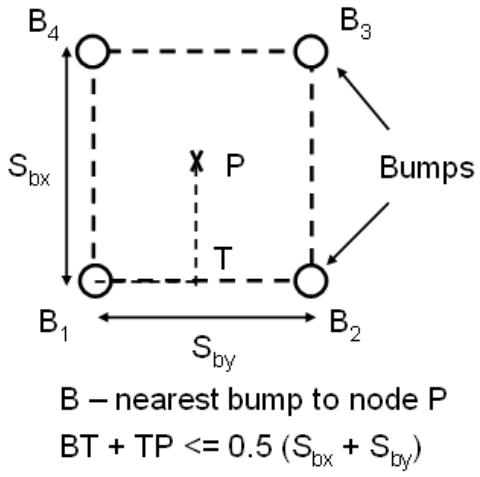



**Figure 48. Calculation of the maximum distance between a node and its nearest power(ground) supply.**

maximum distance between any node P in the area bounded by rectangle $B_1B_2B_3B_4$ to its nearest power (ground) bump (B in Figure 48) obeys the relation

$$\mathrm{BT} + \mathrm{TP} \leq 0.5(S_{bx} + S_{by)}. \tag{37}$$

Then,

$$\max_i \{R_{i-s}\} \leq 0.5\,(S_{bx} + S_{by}) \max_l \left\{R^l_{pul}\right\}, \tag{38}$$

and

$$\max_i \{L_{i-s}\} \leq 0.5\,(S_{bx} + S_{by}) \max_l \left\{L^l_{pul}\right\}, \tag{39}$$

where $R^l_{\mathrm{pul}}$ and $L^l_{\mathrm{pul}}$ are the resistance and inductance per unit length, respectively, of lines in metal layer $l$. The fictitious capacitance to ground is computed for different values of $f_{\max}$ and $\max_i \{L_{i-s}\}$. In this computation, $f_{\max}$ is chosen as it was for the fictitious inductance. The maximum inductance $\max_i \{L_{i-s}\}$ was computed assuming a per-unit-length inductance of 2 uH-m$^{-2}$ and a worst-case distance of 20 mm from any node to the power supply; therefore, $\max_i \{L_{i-s}\} = 40$ nH. Since $R_{i-s} \ll \omega_{\max} L_{i-s}$ usually, the effect of $\max_i \{R_{i-s}\}$ on the fictitious capacitance to ground was ignored. In Figure 49, the variation of the fictitious capacitance to ground

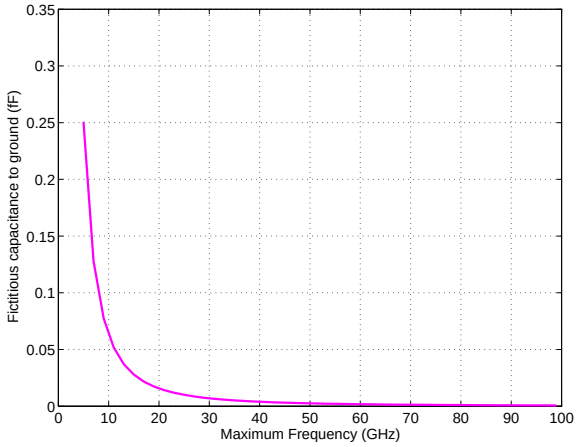


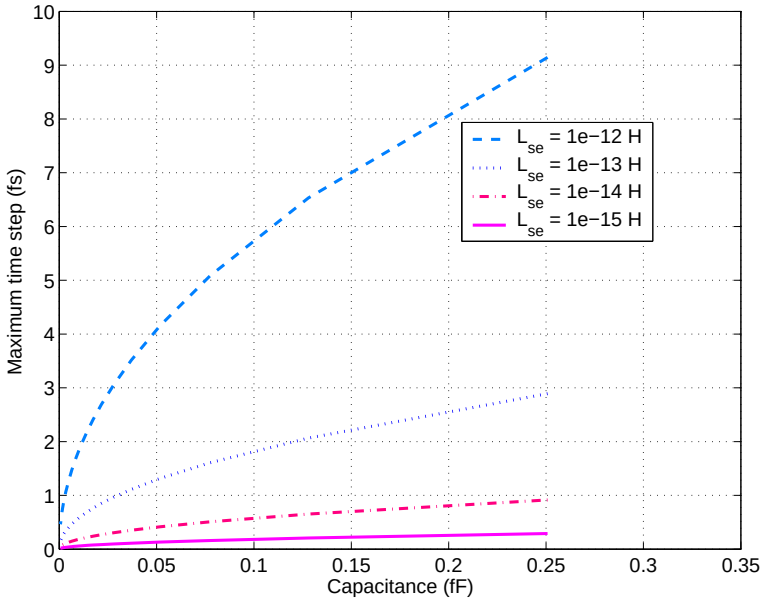

**Figure 49. Variation of fictitious capacitance to ground with maximum frequency of operation for $\max_i \{L_{i-s}\} = 40$ nH and $\max_i \{R_{i-s}\} = 0$.**

with increasing $f_{\max}$ and for $\max_i \{L_{i-s}\} = 40$ nH has been shown. From Figure 49, it can be observed that the fictitious capacitance to ground, $C_{ii}^f$, is less than 0.25 fF when $\max_i \{L_{i-s}\} = 40$ nH. The capacitance $C_{ii}^f$ can be as small as 6 $\times 10^{-19}$ F at 100 GHz when $\max_i \{L_{i-s}\} = 40$ nH.

## 6.7 Computational Complexity of the Transient Simulation

In this section, the memory and the time complexities of the LIM-enabled transient simulation have been derived taking the effect of fictitious elements into account.

Since the transient simulation is based on the LIM, the computational complexity of the transient simulation is same that of the LIM. Since $N_b$ and $N_n$ differ only by a constant factor in the equivalent circuit shown in Figure 42, the memory complexity of the transient simulation is $O(N_n)$, and the time complexity of the transient simulation is $O(N_t N_n)$. Since $\Delta t$ (and hence $N_t$) is independent of $N_n$ from (30), the overall computational complexity of the transient simulation is $O(N_n)$. However, in practice, the time complexity can be more. This is because, though $N_t$ is independent of $N_n$, its value can be comparable to $N_n$, especially with the presence of fictitious latency elements. In the following, it is estimated that the runtime of the LIM-enabled transient simulation for power grid simulation is approximately proportional




(a) $\max_i \{L_{i-s}\} = 40$ nH

**Figure 50. Variation of the maximum time step with the smallest capacitance to ground and with the smallest series inductance.**

to ${N_n}^{2-2.5}$ for $N_n$ on the order of millions.

To find the overall time complexity, the typical range of $N_t$ has to be estimated. Since the total time $T = (N_t - 1)\Delta t$, the worst-case values for $\Delta t$ has to be estimated first. For equivalent circuits such as in Figure 37, the maximum time step, $(\Delta t)_{\max}$, is independent of $N_n$ (see (30)). Since $(\Delta t)_{\max}$ is dependent on the smallest L and C in the circuit from (30) and the smallest L and C are usually the fictitious series inductance and shunt capacitance, the effect of the values of fictitious elements on $(\Delta t)_{\max}$ is studied. For this study, the $(\Delta t)_{\max}$ from (30) was computed for different values of fictitious capacitance to ground shown in Figure 49 and for different values of series inductance of branch shown in Figure 46. When there are no decoupling capacitors, the maximum number of branches connected to a node is only four in the equivalent circuit shown in Figure 37. Therefore, $N_b^i = 4$. In Figure 50, the $(\Delta t)_{\max}$ has been plotted for different values of fictitious capacitance to ground and series inductance of branch. From Figure 50, it can observed that $(\Delta t)_{\max}$ decreases with

the decrease in the smallest fictitious capacitance to ground and with the decrease in the smallest fictitious series inductance of branch. The maximum time step can be as low as 0.01 fs for a series inductance of 1 fH and a fictitious capacitance to ground of $6 \times 10^{-19}$ F, and this scenario happens for $f_{\max} = 100$ GHz and $\max_i \{L_{i-s}\} = 40$ nH from the discussions in Section 6.6.

If a more conservative estimate of $(\Delta t)_{\max} = 1$ fs is considered, and if $T \approx 1$ ns is chosen, then $N_t \geq 1$ million. Therefore, when $N_n$ is on the order of millions, the runtime of the whole transient simulation is approximately proportional to $N_n^2$. For $T = 100$ ns and $(\Delta t)_{\max} = 0.1$ fs, then $N_t = 10^9$. When $N_n$ is on the order of millions, the runtime is proportional to $N_n^{2.5}$. Therefore, for $N_n$ on the order of millions, the total time complexity of the runtime of the overall transient simulation is approximately proportional to ${N_n}^{2-2.5}$. For differential-mode equivalent circuits such as in [7], the capacitance to ground from a node is in couple of femtofarads. In such cases, $N_t \approx 10^6$. For these equivalent circuits, the runtime of the overall transient simulation is approximately proportional to $N_n^2$ for $N_n$ on the order of millions.

### 6.7.1 Remarks

In the proposed transient simulation formulation, the advantages mentioned in Chapter 2 are preserved with a numerical robustness associated with a direct solver-based implicit method. However, unlike the direct solver, the optimal memory complexity is preserved irrespective of the numbering of the nodes.

The drawback of the proposed method is its high time complexity. This high complexity is only because of the small time step of the transient simulation observed in these equivalent circuits. This complexity is expected to be alleviated to $O(N_n)$ using time-step relaxation schemes such as the alternate direction implicit (ADI) methods without compromising the memory complexity. Using an ADI-based method, the dependence of $\Delta t$ on the element values is removed, yielding a $N_t$ that is small compared to $N_n$. Such a fix using ADI methods is common to relax the restriction on the time

step in FDTD method [76], [77]. There have been also efforts to employ ADI-based methods for the on-chip power-grid simulation in mesh-type power grids [43], [38].

## 6.8 Results

In this section, the transient results that demonstrate the accuracy of the LIM-enabled power-grid simulation and the accuracy of the proposed closed-form expressions have been presented. The rest of the section is organized as follows: First, the transient results have been obtained for a small problem. Second, the transient results pertaining to the accuracy of the closed-form expressions have been presented for a large problem.

### 6.8.1 Small Problem

#### *6.8.1.1 Test Setup*

The test setup consists of an on-chip PDN like the one in Figure 36, with three metal layers; M1 is the metal layer closest to substrate, M3 is the metal layer farthest from substrate, and M2 is the metal layer between M1 and M3. In Figure 17 (Chapter 2), the cross-sectional view of this on-chip PDN has been shown. Only 400 um $\times$ 400 um of the region starting from (0, 0) has been considered for this test. The total number of nodes, $N_n$, is 1900. The per-unit-length parameters of the lines in the different layers have been listed in Table 2. In Table 2, the lines in M1 have a capacitance to ground, while the lines in M2 and M3 do not have a capacitance to ground. The via resistances and inductances and the crossover capacitances between different metal layers have been listed in Table 3. The arrangement of the power- and the ground-supply bumps in M3 has been shown in Figure 51. The leakage and the switching power densities have been chosen as 125 mW-mm$^{-2}$ each. The leakage current has been modeled as a DC current source; these sources have been distributed uniformly in M1; each leakage current has an amplitude of 49 uA. The switching current has been modeled as a periodic triangular pulse stream with rise time = 10 ps, fall time

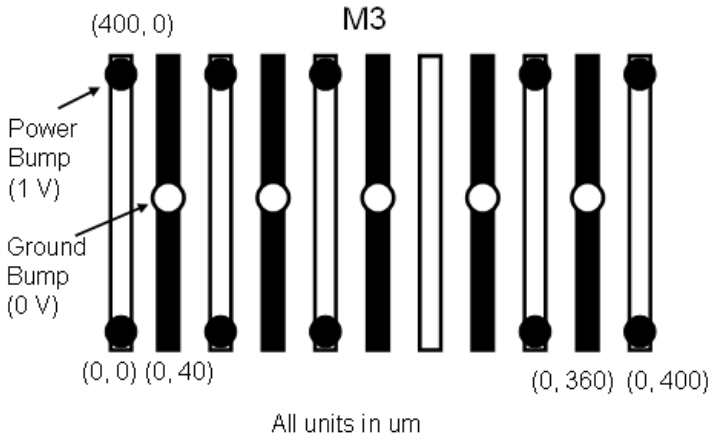



**Figure 51. The arrangement of power- and ground-supply bumps in M3.**

**Table 2. The per-unit-length R, L, C parameters of power-ground lines in different layers of the on-chip PDN**

| Metal layer | R (Ω/m) | L (H/m) | C (F/m) |
|---|---|---|---|
| M1 | 17246.7 | 7.357e-7 | 1.884e-10 |
| M2 | 6750 | 1.3e-6 | 0.0 |
| M3 | 3750 | 1.425e-6 | 0.0 |

= 20 ps, delay time = 0, period = 200 ps, and peak amplitude = 641 uA; all sources in the rectangular area bounded by the locations (x = 50 um, y = 50 um) and (x = 350 um, y = 350 um) have been assumed to be switching starting from $t = 0$. A total simulation time of 300 ps has been chosen. For the DC simulation, the method proposed in Chapter 4 has been used.

**Table 3. Via resistance and inductance and crossover capacitance between different metal layers**

| Metal Lyr. | Via Res. (mΩ) | Via Ind. (pH) | Crossover Cap. (fF) |
|---|---|---|---|
| M1-M2 | 34.5 | 1.47 | 0.4 |
| M2-M3 | 13.5 | 2.6 | 1.63 |

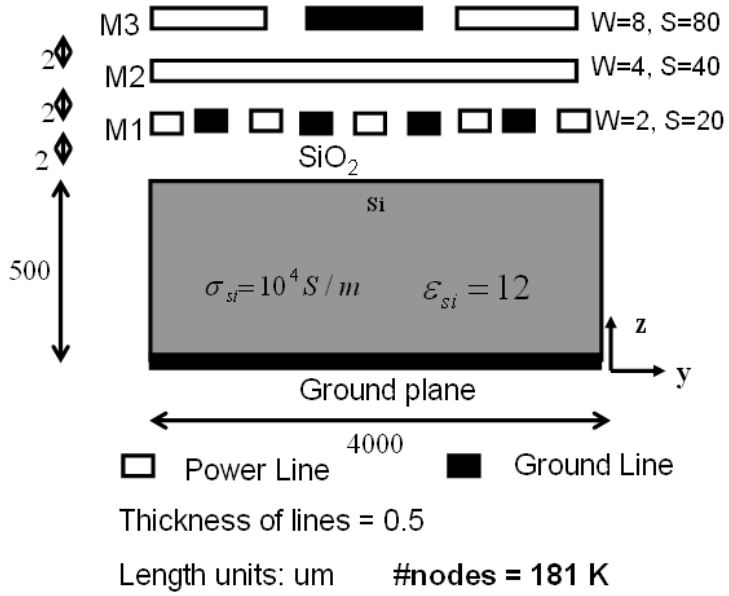



**Figure 52. The cross-sectional view of the on-chip PDN.**

*6.8.1.2 Accuracy of the LIM-enabled Transient Simulation*

The accuracy of the transient results using the LIM is compared with that from HSPICE. To enable the LIM, fictitious capacitance of 1 fF is added to ground from all nodes in M2 and M3, and $K_L = 10^{-3}$ (see (31)) for all crossover capacitances. The differential transient voltages were computed at (x = 200 um, y = 200 um) using both LIM and HSPICE, and these results have been compared in Figure 53. The time step was computed as 17.8 fs. From Figure 53, it can be observed that result from LIM matches well with that from HSPICE. The maximum instantaneous relative error was less than 0.06%. This test demonstrates the accuracy of the LIM.

*6.8.1.3 Accuracy of the Proposed Closed-Form Expressions for Fictitious Elements*

The computation of the fictitious inductance is relatively easy compared to that of the fictitious capacitance to ground. Since all the terms in (32) except the term $k_L$ are known before the simulation, the choice of the value of $K_L$ completes the computation of the fictitious inductance. It has been observed from many simulations that $K_L \leq 10^{-3}$ guarantees an accurate result for all problems. The computation of

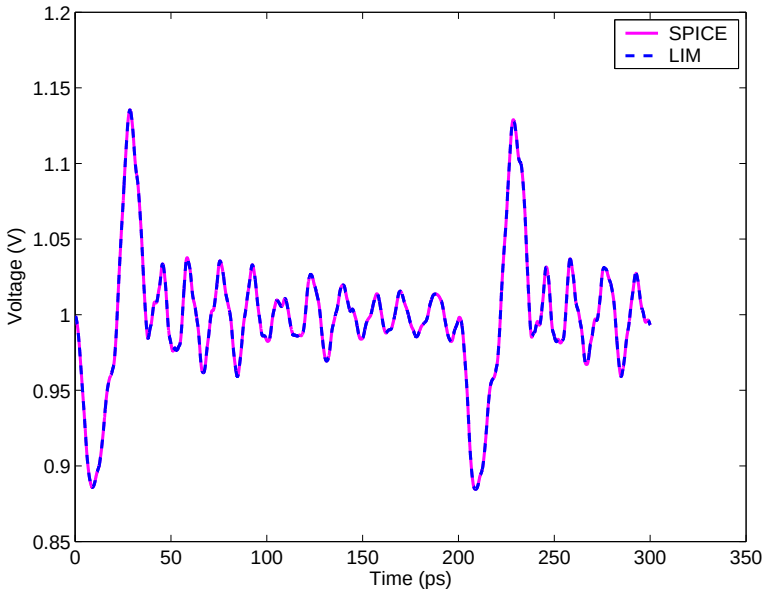



**Figure 53. Comparison of the differential voltage at (x = 200 um, y = 200 um) in M1 from the LIM method with that from HSPICE.**

fictitious capacitance to ground, however, is not straightforward. The term $k_C = 10^{-2}$. From Figure 51, it can observed that the maximum distance between a node and its nearest power supply is less than 280 um. Using (36), (38), and (39), the fictitious capacitance to ground from any node was found out to be 0.0634 fF. Therefore, the capacitance to ground can be chosen any value less than or equal to 0.0634 fF. Using a fictitious capacitance to ground of 0.01 fF ($<$ 0.0634 fF), the differential transient node voltage has been computed at (x = 200 um, y = 200 um). The time step, $\Delta t$, is computed through (30) as 1.78 fs (see Figure 54). It can be noticed in Figure 54 that resistances and inductances connected to the node are not the same in all branches. Therefore, the circuit is inhomogeneous. In Figure 55, the differential node voltage at (x = 200 um, y = 200 um) has been plotted with and without the fictitious capacitance to ground. From Figure 55(a-b), it can be observed that the results are bounded. This demonstrates the accuracy of the upper bound of the time step shown in (30). From Figure 55(a), it can be observed that result with a fictitious capacitance of 0.01 fF agrees well with the result without the fictitious capacitance (this result

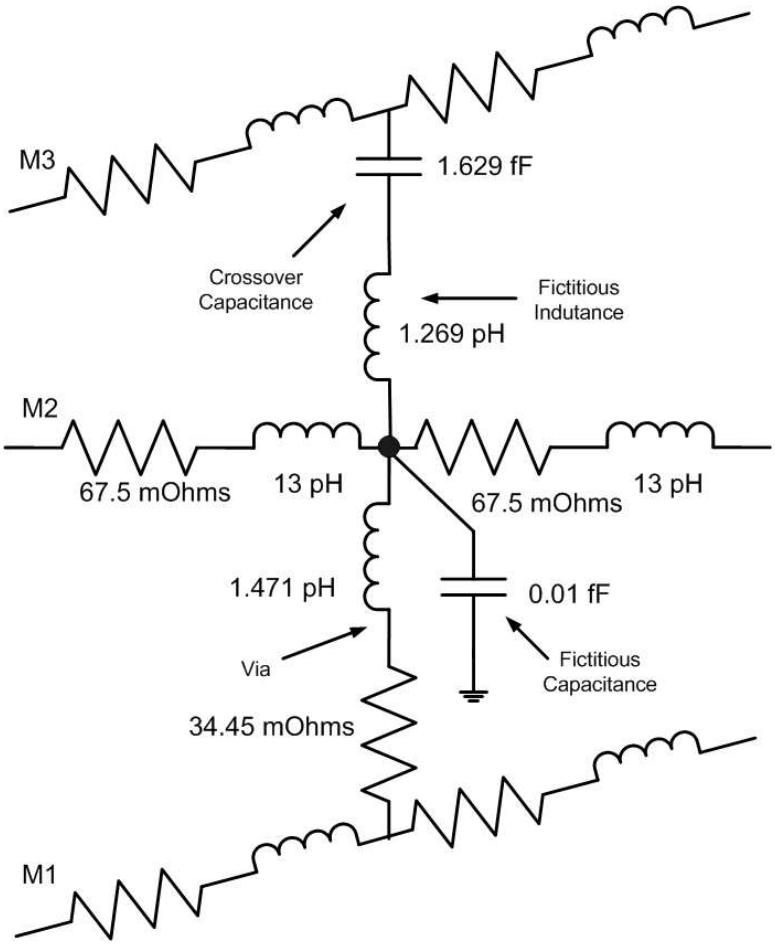



**Figure 54. Time step calculation. Shown is the equivalent circuit near the node that has the smallest time step; $\Delta t = \frac{\sqrt{1.269\text{pH} \cdot 0.01\text{fF}}}{\sqrt{4}} = 1.78$ fs**

.

was obtained using HSPICE). The maximum relative error between the two results is 0.4%. The result in Figure 55(a) demonstrates the accuracy of the closed-form expressions proposed in Section 6.6. From Figure 55(b), it can be observed that the result with a fictitious capacitance to ground of 0.1 fF ($>$ 0.0636 fF) differs from the result without this capacitance. The maximum relative error between the two results is 5.2%. The time step, $\Delta t$, is 5.63 fs when fictitious capacitance to ground is 0.1 fF. Thus, the fictitious capacitance to ground has to be computed carefully if accuracy is not to be compromised.

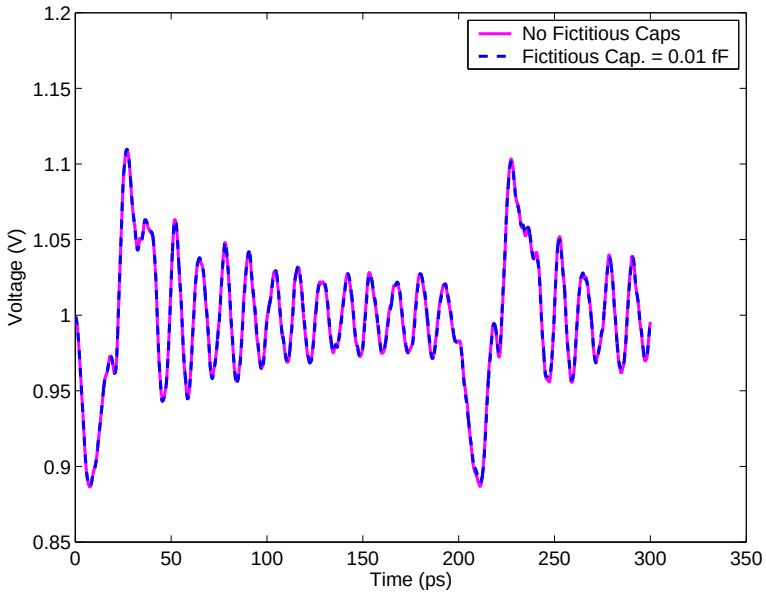
### 6.8.2 Large Problem

#### *6.8.2.1 Test Setup*

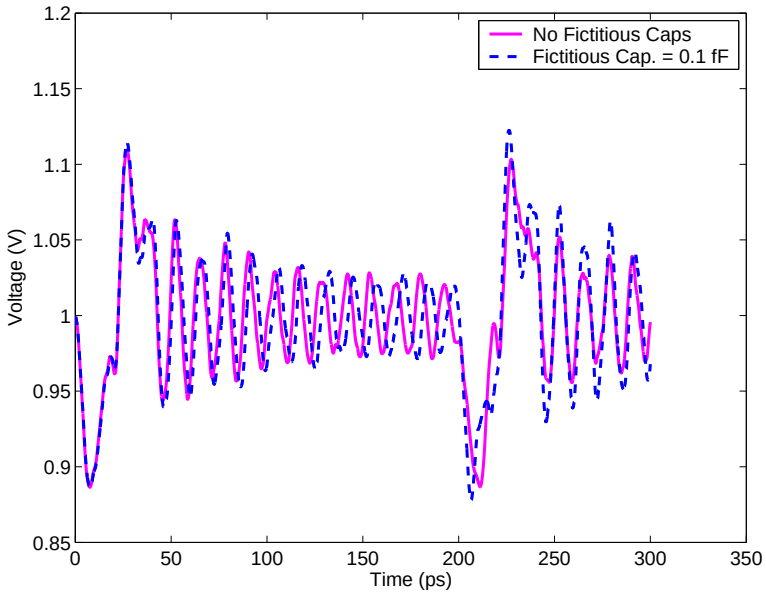
The test setup remains with the following changes: 1) The size of the chip is increased to 4000 um $\times$ 4000 um ($N_n = 180,000$). 2) The new bump locations are as shown in Figure 56. 3) The leakage current sources are distributed in the 4000 um $\times$ 4000 um area in M1. 4) The switching current sources are confined to the center of M1 in the rectangular area bounded by locations (x = 1500 um, y = 1500 um) and (x = 2500 um, y = 2500 um).

#### *6.8.2.2 Accuracy of the Proposed Closed-Form Expressions for Fictitious Elements*

The computation of the fictitious series inductance remains unchanged from that in the small problem (since the crossover capacitance and excitation remain unchanged). However, the fictitious capacitance to ground changes, as the bump locations have changed. From Figure 56, it can be observed that the maximum distance between a node and its nearest power supply is less than 1040 um. Using this distance, and using (36) - (39), the fictitious capacitance to ground is computed to be any value less than 0.033 fF. Since it is a challenge to run this problem in HSPICE, the accuracy of the transient results is shown by observing the convergence of these results with the fictitious capacitance to ground. The results obtained with a fictitious capacitance to ground of 0.01 fF ($<$ 0.033 fF) have been used as the reference result for showing




(a) Fictitious capacitance to ground = 0.01 fF




(b) Fictitious capacitance to ground = 0.1 fF

**Figure 55. Comparison of the transient results obtained with and without the fictitious capacitance to ground.**

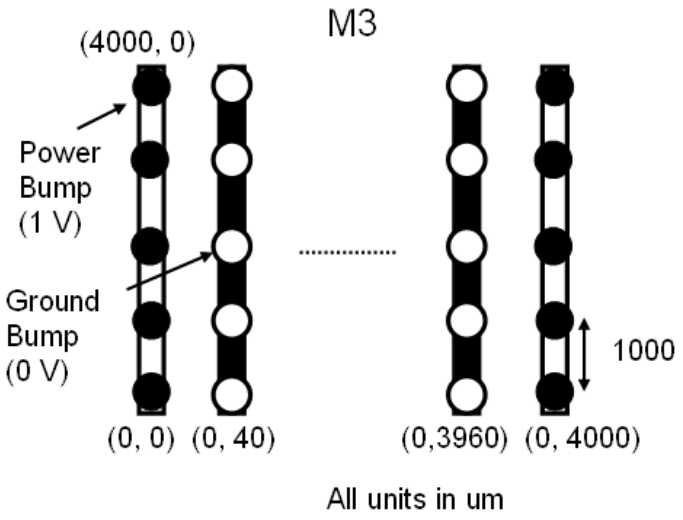
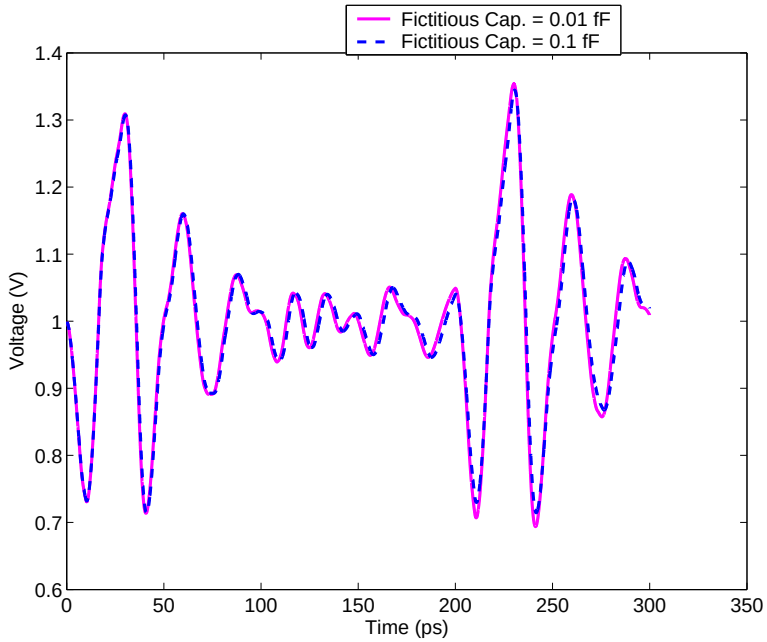



**Figure 56. The new arrangement of power- and ground-supply bumps in M3.**

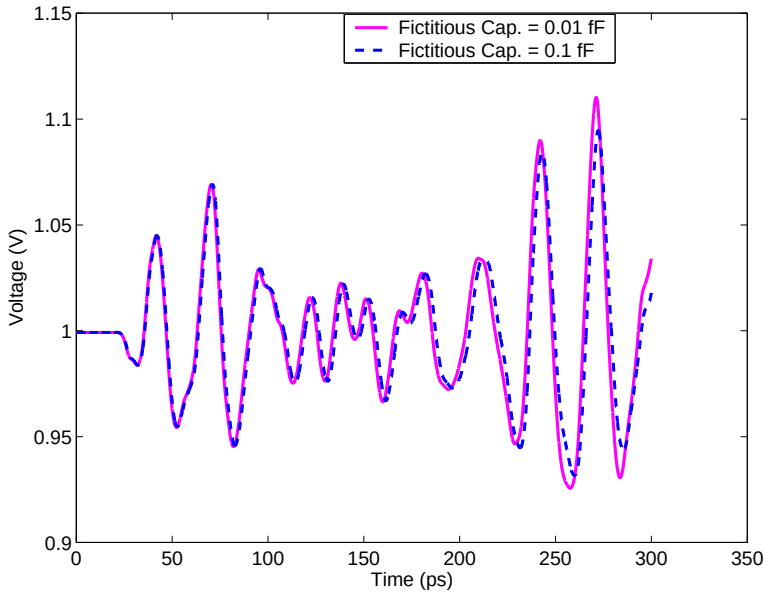
the convergence. In Figure 57, the differential transient voltages with a fictitious capacitance of 0.01 fF are compared with those obtained with a fictitious capacitance of 0.1 fF. The time steps with both these capacitances are same as those in the small problem. From Figure 57, it can be observed that the transient results are almost the same. The maximum relative error between the results obtained with the fictitious capacitance of 0.01 fF and the results obtained with the fictitious capacitance of 0.1 fF is 5.4%. This error is 22% when 1 fF was used and is 36% when 10 fF was used. Therefore, the maximum relative error keeps reducing with the decrease in the fictitious capacitance. Therefore, the transient results with a fictitious capacitance to ground of 0.033 fF would have a maximum relative error of less than 5.4%. Thus, this result also demonstrates the accuracy of the proposed closed-form expressions.

### 6.8.3 Memory and Time Requirements

In this section, the memory and time taken by the simulation for the two problems are described. In Table 4, the time and memory requirements of the proposed method are shown. For the small problem ($N_n$ = 1.9 K), the time taken per time step of the transient simulation is approximately 0.058 s. The memory required is 1.02




(a) Differential transient voltage at (x = 2000 um, y = 2000 um)




(b) Differential transient voltage at (x = 2000 um, y = 600 um)

**Figure 57. Convergence of transient results with the reduction in the fictitious capacitance to ground.**

**Table 4. Time and Memory requirements of the Proposed Transient Simulation Approach**

| $N_n$ | $\Delta t$ | $N_t$ | Time taken per time step | Memory |
|---|---|---|---|---|
| 1.9 K | 1.78 fs | 168 K | 0.058 s | 1.02 MB |
| 181 K | 1.78 fs | 168 K | 5.8 s | 92.7 MB |

MB, which includes the memory required for storing the geometry (0.76 MB) and the memory required for the DC solution (0.26 MB). For the transient solution, no additional memory is required, as the node voltages (and branch currents) are solved independently. For the large problem ($N_n$ = 181 K), the time taken per time step of the transient simulation is 5.8 s. The memory required is 92.7 MB (= 68.35 MB for geometry + 24.35 MB for DC solution). It can be observed that both the memory requirement and the time taken per time step of the transient simulation scale linearly with the problem size, $N_n$ and, therefore, are optimal in complexity. The total time taken for the whole transient simulation is affected by the value of $N_t$. Since $\Delta t = 1.78$ fs and the total simulation time is 0.300 ns, $N_t \approx 168$ K. Such a large $N_t$ increases the total simulation time. However, $\Delta t$ (and therefore $N_t$) is independent of $N_n$, as $\Delta t$ depends only on the smallest $L$ and $C$ values. Therefore, the proposed method is advantageous in terms of overall run time when $N_n \gg N_t$. Such a situation arises either when $T$ is small and/or $\Delta t$ is large or when $N_n$ is large.

In HSPICE, the memory required for the small and the large problem are 21.76 MB and > 1.5 GB, respectively. While the small problem was completed faster than the proposed method (by two times for the same $\Delta t$), the large problem was not completed because of the large memory requirements. This high memory requirement in HSPICE is primarily because of the memory requirements of the direct solver in HSPICE. Since the memory and time requirements of a direct solver depends on the way the nodes are numbered, these requirements can be improved with careful node numbering. One of the better complexity with a direct solver comes with a

nested dissection node ordering [50]. For problems arising out of discretizing partial differential equations in regular 2-D grid, it has been shown [50] that ordering the nodes in a nested dissection manner makes the memory complexity $O(N_n \log_2 \sqrt{N_n})$ and the time complexity $O(N_n^{\frac{3}{2}})$, where $N_n = M^2$, and $M$ is the total number of nodes in one dimension. For a 3-D grid, the corresponding complexities are shown to be $O(N_n^{\frac{4}{3}})$ and $O(N_n^2)$, respectively, where $N_n = M^3$ [78]. In a power grid, the number of nodes in a line along the direction of the height of the chip is usually a constant ($\leq$ number of metal layers). Therefore, for power grids $M^2 \leq N_n \leq M^3$, where $M$ is number of nodes in a single power/ground line. Therefore, one of the better memory and time complexities achievable for a power grid problem can be in between the complexities of the nested dissection-based direct solvers in two and three dimensions. However, the proposed method guarantees $O(N_n)$ memory complexity and $O(N_n)$ time complexity per time step for the power grid problem independent of the way the nodes are numbered. Moreover, the proposed method is as robust as a direct solver in terms of accuracy and convergence. Also, $N_t$ is independent of $N_n$. Therefore, for $N_t \ll N_n$, the proposed method may also be advantageous in terms of runtime.

## 6.9 Summary

The on-chip power-grid simulation has been performed using the LIM in the equivalent circuits of on-chip PDNs in which some of the nodes did not have a capacitance to ideal ground and some of the nodes had a floating capacitance between them. A small capacitance to ground was added to those nodes that did not have this capacitance, and a small series inductance was added to those capacitive branches that did not have this inductance. The closed-form expressions for the fictitious capacitance to ground and fictitious series inductance have been proposed. The accuracy of the LIM-enabled power-grid simulation has been shown. The accuracy of the proposed

closed-form expressions has been demonstrated. It has been shown that the memory complexity for the overall transient simulation is $O(N_n)$. It has been shown that the time complexity per time step of the transient simulation is $O(N_n)$. It has been found that because of the small values of the fictitious elements, the maximum time step of the transient simulation becomes small and therefore the time requirements for the overall transient simulation becomes high. It has been estimated that the runtime of the overall LIM-enabled transient simulation is approximately proportional to $N_n^{2-2.5}$ for problem sizes on the order of millions.

# CHAPTER 7

# ON-CHIP LIM INCLUDING ON-CHIP DECOUPLING CAPACITANCE AND PACKAGE PDN EFFECTS

## 7.1 Introduction

In Chapter 6, the LIM was proposed as an efficient method for simulating PSN in on-chip power grids even in the presence of branch capacitors. However, in Chapter 6, the performance of LIM was demonstrated only with crossover capacitors. In this chapter, the performance of LIM is demonstrated with even on-chip decoupling capacitors. On-chip decoupling capacitors place a bigger strain on the simulation than crossover capacitors do. Unlike previous chapters, the nonideal nature of the C4 bumps and package PDN is also considered in this chapter. The need for modeling this nonideality is demonstrated, and LIM is extended to model this nonideality. Finally, with the new equivalent circuit and LIM formulation, the effect of the on-chip inductance on the PSN is studied. From the study, it is concluded that on-chip inductance does not affect the PSN significantly. However, it can induce sudden transient spikes in the power-supply voltage. *The contributions of this chapter are the demonstration of LIM performance with on-chip decoupling capacitors and nonideal model for C4 bumps and package PDN and the demonstration of the effects of the on-chip power-grid inductance on PSN.*

## 7.2 Background and Prior Work

On-chip decoupling capacitors are essential to controlling PSN [25], [22], [23]. Modern digital microprocessors allocate close to 10% of the chip's area for decoupling capacitors (see [13]). These capacitors are usually thin-oxide capacitors and are explicitly added to the chip. Besides this capacitance, there are some intrinsic decoupling capacitances too in the chip. The nonswitching circuit capacitance and the N-well

capacitance of CMOS circuits are examples of intrinsic decoupling capacitance. A series RC circuit is usually employed to model this capacitance.

It was described in Chapter 1 that developing computational efficient techniques for on-chip power grid simulation has been the primary objective of most of the prior work in this area. To efficiently simulate large problems, only simplified equivalent circuits for different parts of the on-chip PDN are employed. However, there has not been much effort on how the simplifications affect the accuracy of PSN simulation.

One of the simplification is to ignore the effect of the package, by assuming ideal DC voltage sources at C4 bumps, while simulating the PSN in the chip. Most of the prior work has advocated the need for including the effect of package while simulating the chip. However, no work to the author's knowledge has clearly demonstrated this need.

Another common simplification is to ignore the effect of the inductance of on-chip power grids, by assuming only a distributed RC model instead of a RLC model for power grids. In [79], [80], this assumption was justified using a PEEC-based equivalent circuit for power grids. PEEC-based equivalent circuits guarantee good accuracy but are not preferred in on-chip power grid simulation in the early stages of the grid design. This nonpreference has to do with the computational inefficiency of power grid simulators when dealing with PEEC-based circuits. It is of interest to pursue this study in simplified on-chip power grid equivalent circuits, such as the one shown Figure 37. However, this study has not been done.

Similarly, most of the prior work for on-chip power grid simulation does not model crossover capacitance. Some of the prior work [45], [31], [42], Chapter 3 of this dissertation model this capacitance. Among these, references [45], [31], [42] do not clearly demonstrate its need in on-chip PSN simulation, especially when on-chip decoupling capacitances are also present. In Chapter 3, on-chip PDN equivalent circuits employed did not consider on-chip decoupling capacitance and the nonideal nature of

package PDN. Therefore, the effect of the crossover capacitance on PSN has yet to be demonstrated.

In this chapter, performance of LIM is demonstrated in presence of on-chip decoupling capacitors. The need for modeling the nonideality of C4 bumps and the package PDN are demonstrated. The effect of the on-chip power grid inductance on the PSN simulation is studied using the simplified equivalent circuit shown in Figure 37. From the study, it is concluded that on-chip inductance does not affect the PSN significantly. However, it can induce sudden transient spikes in the power-supply voltage. Finally, the effect of the crossover capacitance on the PSN is studied. It has been observed that crossover capacitances act as decoupling capacitors, confirming the prior-held belief (see Chapter 3) for the first time. However, because of their small values, it has been observed that they do not affect the PSN much.

The rest of this chapter is organized as follows. In Section 7.3, details of modeling on-chip decoupling capacitors using LIM are described. In Section 7.4, these details are presented for the nonideal model for C4 bumps and the package PDN. In Section 7.5, numerical results demonstrating the need for a nonideal model for C4 + package PDN in PSN simulations, the accuracy of LIM transient simulation, and the effect of the on-chip inductance on PSN are presented. Also in this section, the effect of the crossover capacitance on the PSN is simulated. Finally in Section 7.6, the conclusions of this chapter are summarized.

## 7.3 LIM and On-Chip Decoupling Capacitor Modeling

The changes to the on-chip equivalent circuit described in Chapter 6 are the on-chip decoupling capacitors and the nonideal model for C4 bumps and the package PDN. Each on-chip decoupling capacitor is modeled like any branch capacitor would be in an LIM formulation. Therefore, a series RC circuit is used for on-chip decoupling capacitors, and fictitious series inductances are inserted in branch capacitances to

enable LIM.

Unlike other branch capacitances like crossover capacitances, on-chip decoupling capacitance place a greater stress on practical computational complexity of the transient simulation. This stress has to do with the large values of on-chip decoupling capacitances. In Chapter 6, it was shown that, for not compromising the accuracy much, the larger the branch capacitance becomes, the smaller the fictitious inductance (inserted to the capacitance) should get. For the stability of the transient simulation, the time step of the transient simulation is proportional to the square root of the series inductance. As on-chip decoupling capacitances are usually much larger than crossover capacitances, the time step with the former is usually smaller than that only with the latter. As a result of the reduced time step with on-chip decoupling capacitances, the total execution time for the transient simulation increases.

Like other branch capacitances, DC simulation is not affected with a series RC model for on-chip decoupling capacitors, as the capacitance acts as an open circuit at DC.

## 7.4 LIM and C4 + Package PDN Modeling

C4 bumps and package PDN are modeled by a series RL circuit. Such a circuit is placed at locations of C4 bumps. The power and ground supplies are connected to the other end of this RL circuit. The first end of each RL circuit is connected to a power/ground line in the layer of the chip closest to the package. The resistance and inductance of this RL circuit are chosen to reflect the nonideal nature of C4 bumps and the package PDN. The series RL modeling (to account for the effect of C4 bumps and the package PDN) helps to apply LIM without increasing its computational complexity much.

The RL modeling of C4 bumps and package PDN has the following effects on LIM transient simulation. It was shown in Chapter 6 that for not compromising

the accuracy, the fictitious capacitance to ground added to a node (in which it was missing before) is inversely proportional to the total effective series path inductance from this node to the DC power supply. As the inductance of each of this RL circuit is usually much larger than the grid inductance, the fictitious capacitance to ground from a node in the new simulation decreases.

Because of the nonzero resistance in the RL model, DC simulation would be affected. Therefore, the DC simulation should include the resistance of the RL model of C4 bump and the package PDN.

## 7.5 Results

In this section, numerical results demonstrating 1) the need for a nonideal model for C4 + package PDN in PSN simulations, 2) the accuracy of LIM transient simulation, 3) the effect of the on-chip inductance on PSN, and 4) the effect of the crossover capacitance on PSN are presented.

### 7.5.1 Test Setup

The test setup is same as the one described in Section 6.8.1.1 with the following changes: 1) A new arrangement for the power and ground C4 bumps is used (see Figure 58). 2) Decoupling capacitors are placed in M1. 3) The effect of the nonideal nature of C4 bumps and the package PDN are modeled.

The decoupling capacitors are placed as follows (see Figure 59): 1) A total of 22 capacitors (totaling a net capacitance of 40 pF) are ditributed randomly in M1. These capacitors are thought to emulate extrinsic decoupling capacitors (e.g., thin-oxide decoupling capacitors). 2) A total of 441 capacitors (totaling a net capacitance of 4 pF) are distributed uniformly in M1. These capacitors are thought to emulate intrinsic decoupling capacitors (nonswitching circuit + N-well capacitors).

The resistance and inductance of the RL circuit of a C4 bump are assumed as 10 m$\Omega$ and 0.325 nH, respectively. The new time step is computed as 0.935 fs.

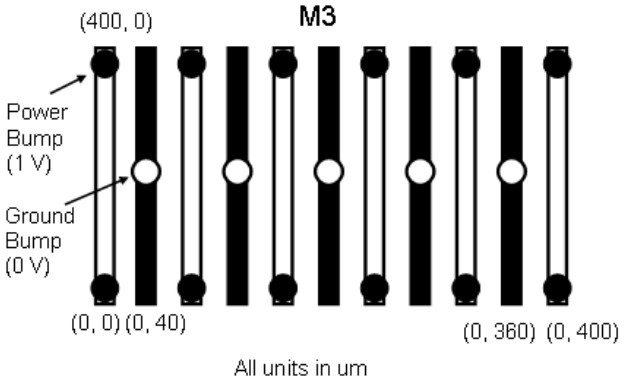



**Figure 58. The arrangement of power- and ground-supply bumps in M3.**

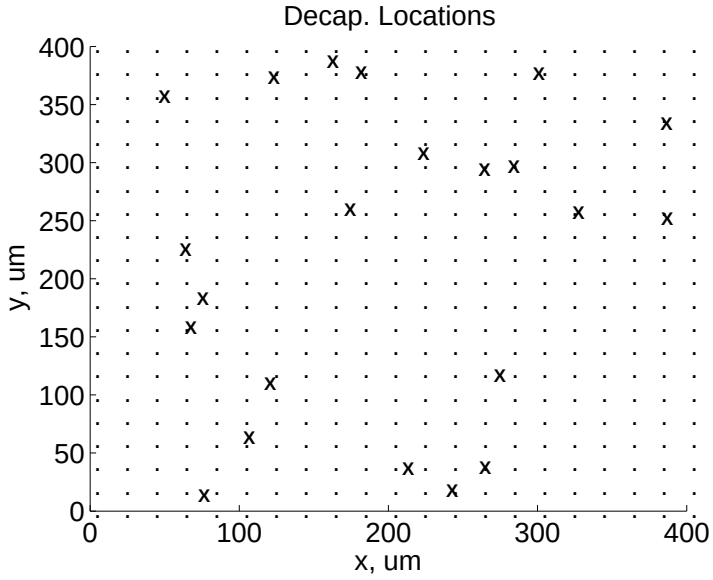
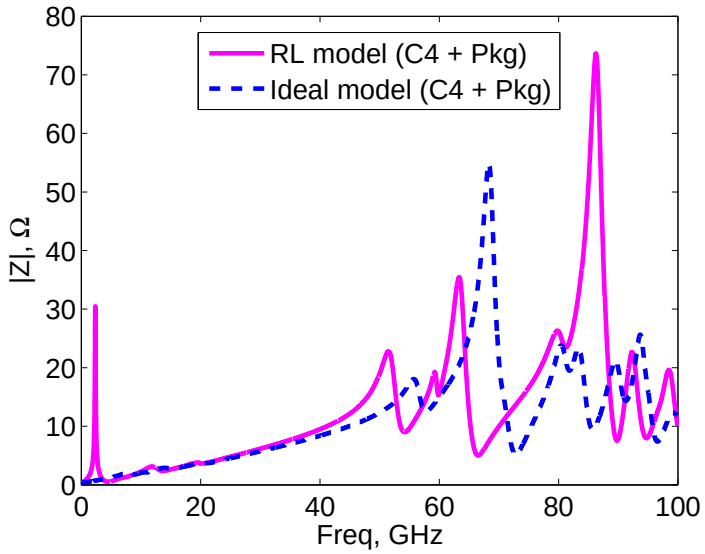


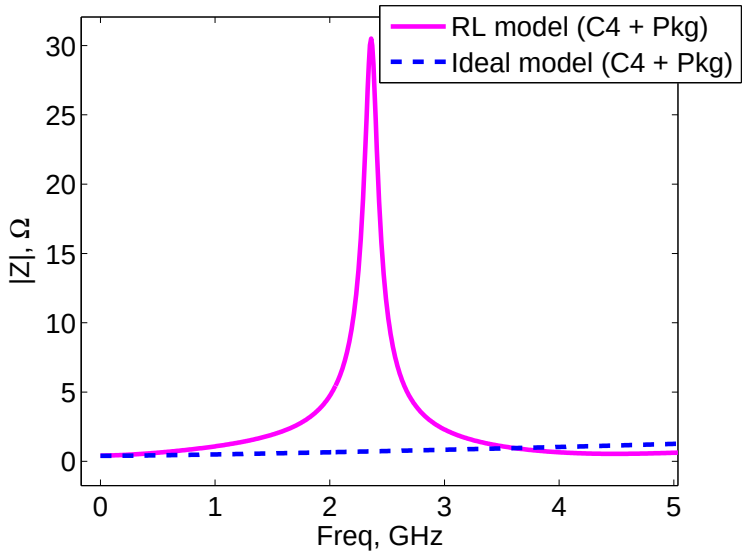

**Figure 59. Placement of extrinsic (denoted as 'x') and intrinsic (denoted as '.') decoupling capacitors in M1**

### 7.5.2 Demonstration of the need for modeling the nonideal nature of C4 bumps and Package PDN

The effect of modeling C4 bumps and the package PDN on the PSN is demonstrated first. The effect is first shown in the frequency domain, and HSPICE is employed for the frequency-domain simulation. A 1V sinusoidal current source is placed at the center of M1. This source is placed between the power node at the center and the ground node closest to this power node. The voltage across the terminals of the current source is computed first with an ideal model for the C4+package and then with a nonideal model for the C4+package. The magnitude of the voltages from these two different models is compared in Figure 60. This voltage is also the input impedance observed at the terminals of the current source. Ideally, if C4 and package parasitics do not matter, then the input impedance (at any location P in M1) obtained with an ideal model for the C4+package should not be much different from the input impedance (at the location P in M1) obtained with a nonideal model for the C4+package. However from Figure 60, it can be observed that there is a significant difference between the input impedances. The input impedance with a RL model for the C4 and package has a resonance near 2.5 GHz. This resonance is actually the chip-package resonance, which is usually observed at a much lower frequency (near 0.2-1 GHz). The high resonant frequency observed in Figure 60 is primarily because of the small area (400 um $\times$ 400 um) of the chip considered. This high resonant frequency would decrease with the increase in the decoupling capacitance and with the increase in the area of the chip. The input impedance with an ideal model for the C4 and package does not have such a resonance (see Figure 60(b)). Note that the inductance of the RL model is a based on a conservative estimate. Therefore, it is clear from Figure 60(b) that an ideal model for the C4+package may not be acceptable and some kind of model capturing the nonideality of the C4+package is necessary.




(a) Input impedance at the center of M1




(b) Zoomed-in version of the input impedance near 1-5 GHz

**Figure 60. Comparison of the input impedance at the center of M1 with an ideal model and a nonideal RL model for C4+package.**

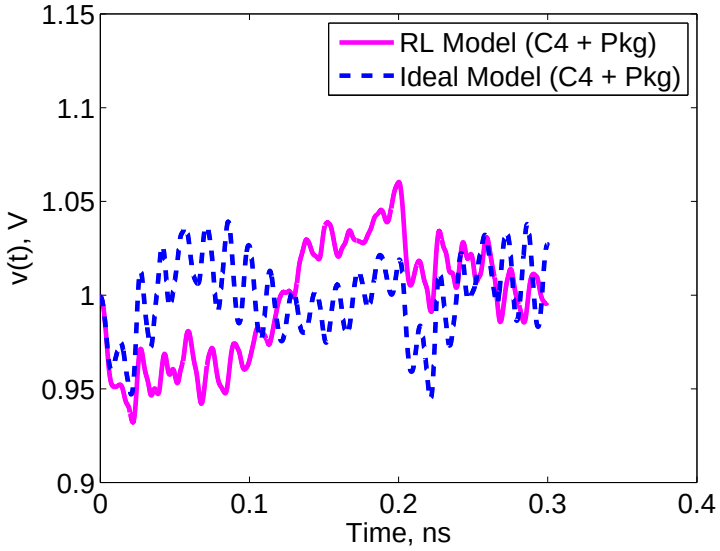
This disparity can also be observed in the time domain. For this purpose, the switching current sources have to be employed. For this test, the switching and leakage currents are excited in the same way (i.e., same specification and distribution) as they were in Chapter 6. Differential transient voltage is computed at the center (x = 200 um, y = 200 um) of M1 and at (x = 200 um, y = 60 um) using HSPICE. The HSPICE results with an ideal model for the C4 + package and with a RL model for the C4 + package are compared in Figure 61. The large disparity between the two results once again demonstrate the need for modeling the nonideal nature of C4 + package.

### 7.5.3 Accuracy of LIM Formulation

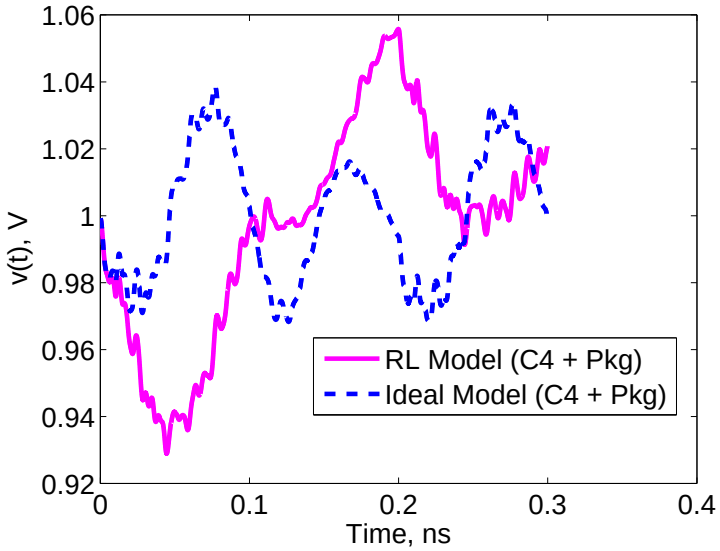
Next, the accuracy of the LIM formulation (for the transient simulation) is demonstrated. For this purpose, the transient results Figure 61 are computed using LIM, and the LIM results are compared with the corresponding HSPICE results. It is to be noted that the fictitious elements are present only in the LIM simulation. In Figure 62, the LIM and HSPICE results obtained with an ideal model for the C4 + package are compared. The close agreement between the LIM and HSPICE results demonstrates the accuracy of the LIM in the presence of decoupling capacitors. This accuracy demonstration is repeated this time however with a RL model for the C4+package. In Figure 62, the corresponding LIM and SPICE results are compared. The close agreement once again between the LIM and HSPICE results demonstrates the accuracy of the LIM in the presence of a RL model for the C4+package.

### 7.5.4 Effect of On-Chip Decoupling Capacitance on PSN

The effect of the on-chip decoupling capacitance on the PSN is demonstrated next. It is common knowledge that on-chip decoupling capacitors help reduce the high-frequency components of the PSN. Therefore, with an increase in the value of decoupling capacitance, a decrease in the power-supply fluctuation is expected. This effect

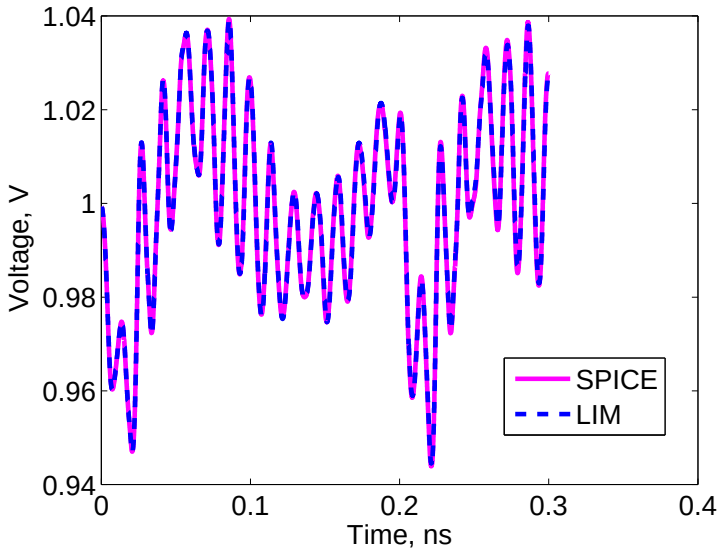



(a) Differential power-supply voltage at (x = 200 um, y = 200 um) in M1

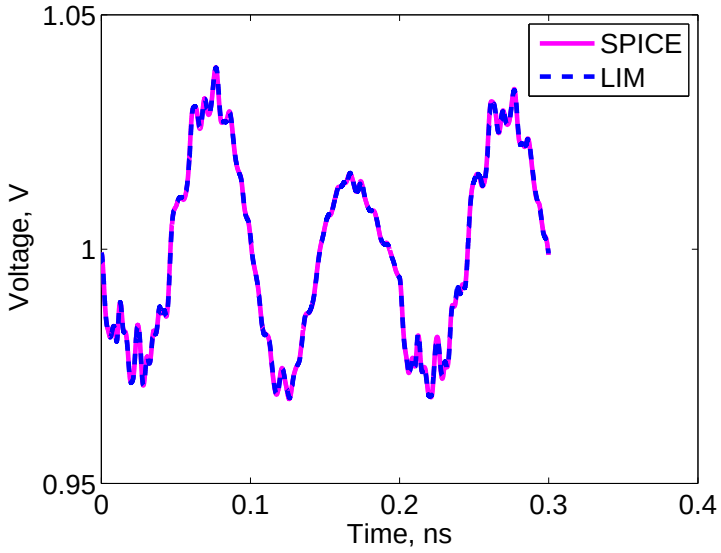



(b) Differential power-supply voltage at (x = 200 um, y = 60 um) in M1

**Figure 61. Comparison of the HSPICE transient results obtained with an ideal model and with a RL model for the C4 + Package.**

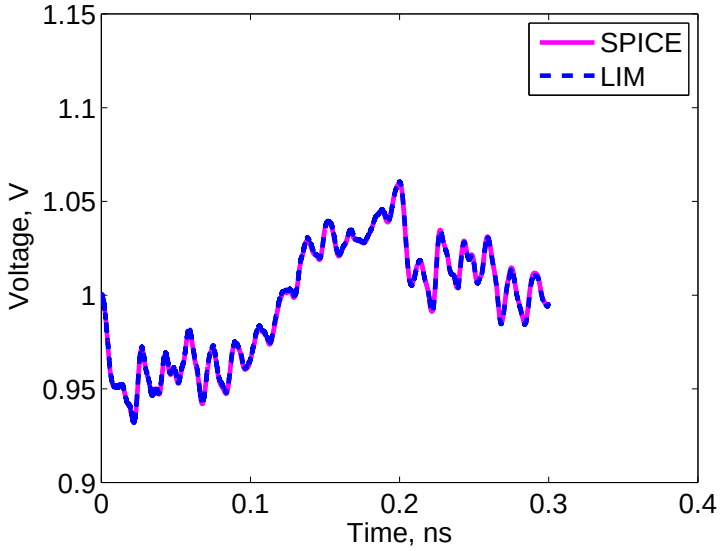



(a) Differential power-supply voltage at (x = 200 um, y = 200 um) in M1

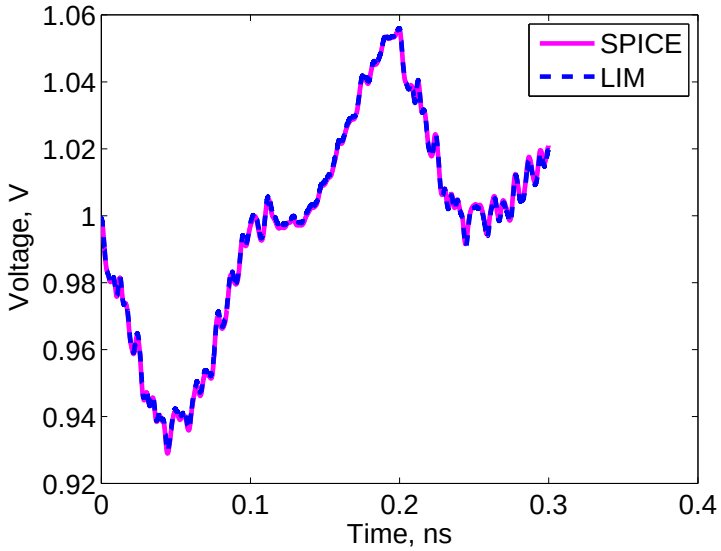



(b) Differential power-supply voltage at (x = 200 um, y = 60 um) in M1

**Figure 62. Comparison of the transient results from LIM and SPICE with an ideal model for the C4+package.**




(a) Differential power-supply voltage at (x = 200 um, y = 200 um) in M1



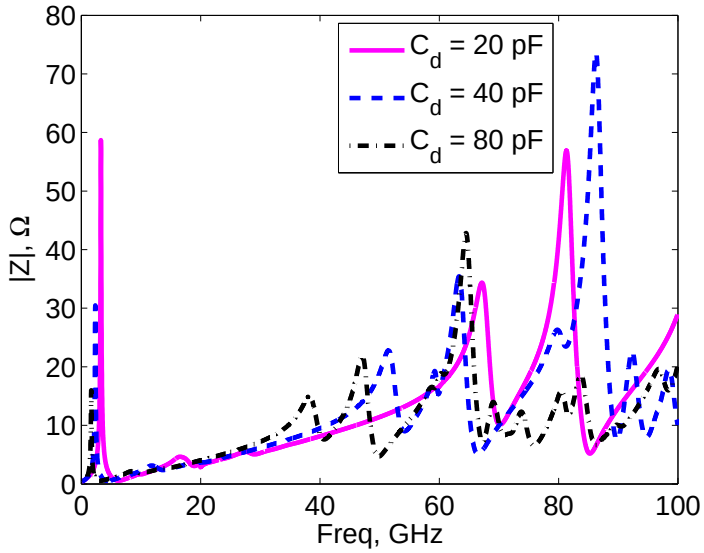

(b) Differential power-supply voltage at (x = 200 um, y = 60 um) in M1

**Figure 63. Comparison of the transient results from LIM and SPICE with a RL model for the C4+package.**

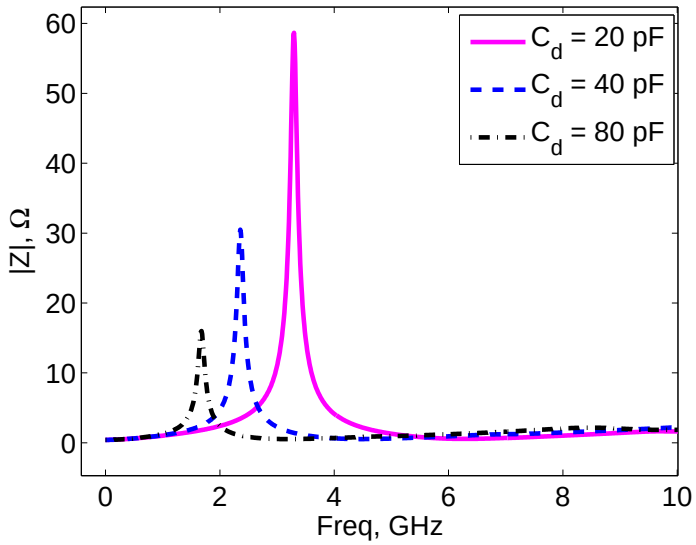
is demonstrated. The results shown in Figure 63 are obtained with two more values (one greater than 40 pF and one lesser than 40 pF) of decoupling capacitance. The locations of decoupling capacitors in all the three cases are retained the same. The input impedance is computed (as before) at the center of M1 with decoupling capacitances of 20 pF, 40 pF, and 80 pF. The resulting input impedances are compared in Figure 64. From Figure 64(b), it can be noticed that the peak amplitude of the input impedance decreases with the increase in the value of the decoupling capacitance. As the magnitude of the power-supply voltage fluctuation at a location is directly proportional to the magnitude of input impedance observed the same location, the decrease in magnitude of the input impedance directly corresponds to a reduction in the peak magnitude of the PSN. From Figure 64(b), it can also noticed that the resonant frequency decreases with the increase in the value of the decoupling capacitance. Capturing this behavior of the decoupling capacitance demonstrates the accuracy of the equivalent circuit models proposed and employed in this dissertation. This effect is now demonstrated in the time domain. In Figure 65, the power-supply voltages obtained with three different values of decoupling capacitances are compared. From Figure 65, it can be observed that with the increase in the value of the decoupling capacitance, the peak amplitude of the transient power-supply voltage decreases in amplitude.

### 7.5.5 Effect of Grid Inductance on PSN

The effect of the inductance of power-ground lines in the on-chip PDN on the PSN is studied next. The input impedance is computed (as before) with and without the (power-ground) grid inductance. The test set up is same as the one described early in this section. In Figure 66, the input impedances with and without grid inductances are compared. It can be noticed from Figure 66(a), the input impedances differ significantly at high frequencies ($> 10$ GHz), thereby making it necessary to understand

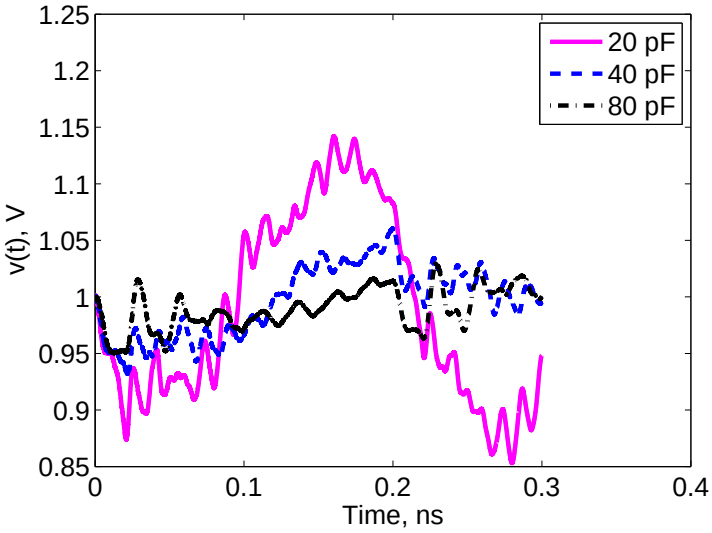



(a) Input impedance at the center of M1

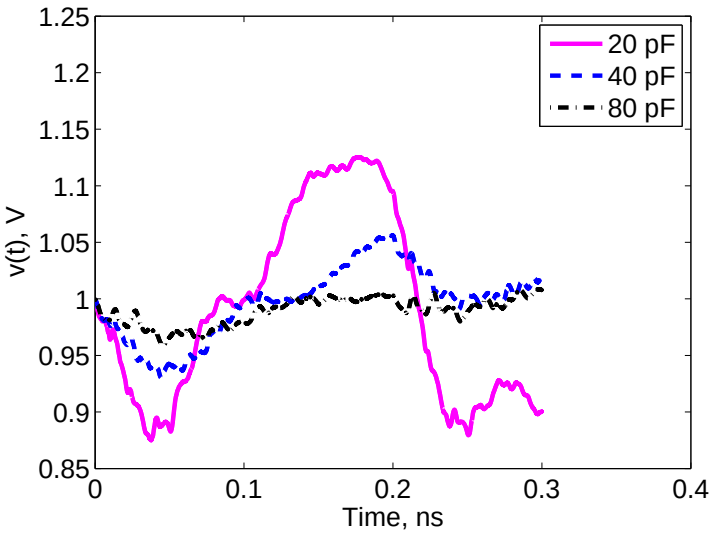



(b) Zoomed-in version of the input impedance near the chip-package resonance

**Figure 64. Comparison of the input impedances obtained with decoupling capacitances of 20 pF, 40 pF, and 80 pF. The capacitance in figure denotes only the total extrinsic capacitance. A total intrinsic capacitance of 10% of the extrinsic capacitance is also included.**




(a) Differential power-supply voltage at (x = 200 um, y = 200 um) in M1




(b) Differential power-supply voltage at (x = 200 um, y = 60 um) in M1

**Figure 65. Comparison of the differential power-supply voltage for three different values of the total decoupling capacitance. The capacitance in figure denotes only the total extrinsic capacitance. A total intrinsic capacitance of 10% of the extrinsic capacitance is also included.**

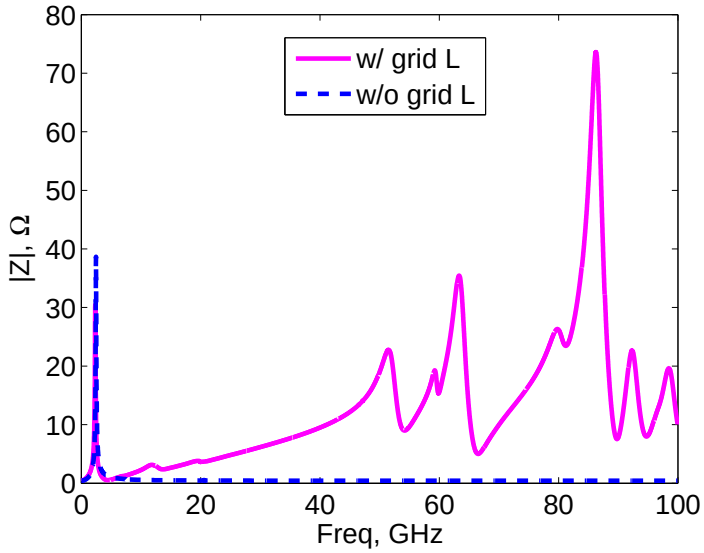
this difference and its effect on PSN. The input impedance without the grid inductance exhibits a resonance at a frequency near 2.46 GHz, determined by the package inductance and the on-chip capacitance. At frequencies greater than 2.46 GHz, the input impedance without the grid inductance is dominated by the reactance of the decoupling capacitance, which is inversely proportional to the frequency. Therefore, the magnitude of the input impedance without the grid inductance effectively decreases with frequency for frequencies greater than the chip-package resonant frequency. This behavior of the input impedance without the grid inductance is consistent with the existing knowledge in the literature.

Unlike the input impedance without the grid inductance, the input impedance with the grid inductance highlights the missing parts of the existing knowledge. The input impedance with the grid inductance differs from the input impedance without the grid inductance in the following ways: 1) The chip-package resonant frequency decreases with the inclusion of the grid inductance (see Figure 66(b)). The resonant frequency (loosely defined as the frequency at which the magnitude of the input impedance peaks) with the grid inductance is 2.36 GHz. The resonant frequency without the grid inductance is 4.4% more than the resonant frequency with the grid inductance. This reduction in the resonant frequency is intuitively consistent with the increase in the value of the inductance. 2) The peak amplitude of the input impedance near the chip-package resonance decreases with the inclusion of the grid inductance. There is a 27% difference between the peak amplitudes, suggesting the need to not ignore the grid inductance when simulating the PSN. 3) There are extra resonances in the input impedance with the grid inductance for frequencies greater than the chip-package resonant frequency. Some of these resonances have amplitudes larger than the amplitude at the chip-package resonant frequency, necessitating a study of these resonances on the PSN in the time-domain. The extra resonances

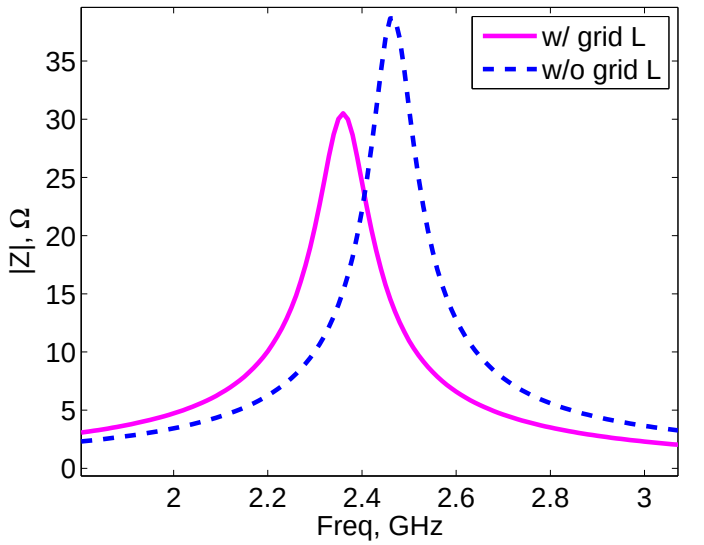
because of the grid inductance can introduce sudden transient spikes to the power-supply voltage. The effect of the grid inductance is now studied in the time domain. The switching and leakage current source specifications are same as before. The differential transient power-supply voltages are computed with and without the grid inductance and are compared in Figure 67. It can be noticed from Figure 67 that the power-supply voltages with and without the grid inductance are the same in an average sense (i.e., when the spikes are averaged). The sudden spikes in the power-supply voltage with the grid inductance are because of the extra resonances observed with the grid inductance (see Figure 66). The extra spikes in the power-supply voltage with the grid inductance can make the fluctuation in the power-supply voltage not acceptable. This effect is demonstrated next. The transient simulation performed just above is repeated with a decoupling capacitance of 80 pF. The differential transient power-supply voltage is computed this time at (x = 200 um, y = 180 um) with and without the grid inductance. These voltages are compared in Figure 68. When a 5% threshold is allowed for the power-supply voltage fluctuation, then the power-supply voltage with the grid inductance alone is more than this threshold (see Figure 68(b)) temporarily.

### 7.5.6 Effect of Crossover Capacitance on PSN

The effect of the crossover capacitance on the PSN is demonstrated next, first in the frequency domain and then in the time domain. For this test, a decoupling capacitance of 40 pF is used, and the grid inductances are included. The input impedance is computed at the center of M1 as before with and without the crossover capacitance. The resulting impedances are compared in Figure 69. From Figure 69, it can be observed that the crossover capacitance does not significantly affect the input impedance. Though there are some perceptible differences at high frequencies, these differences may not affect the transient results much, as the energy in switching

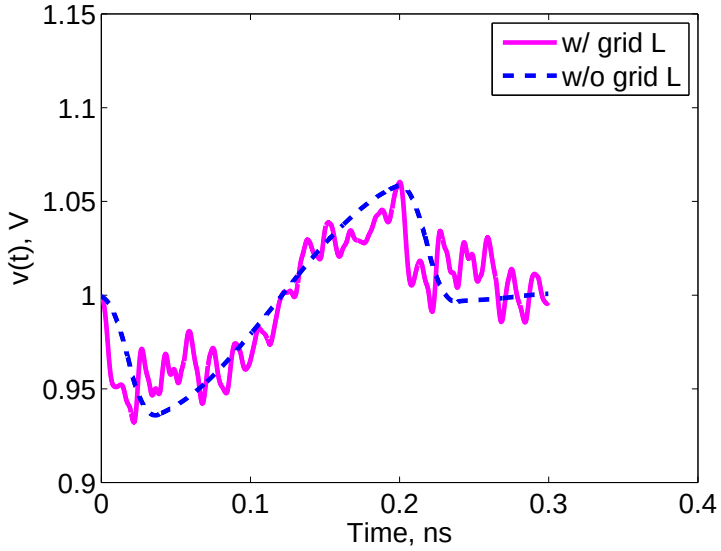



(a) Input impedance at the center of M1

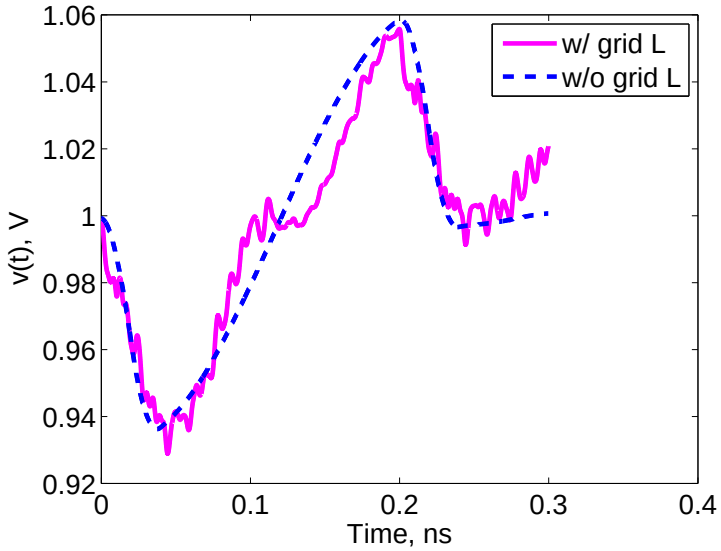



(b) Zoomed-in version of the input impedance near the chip-package resonance

**Figure 66. Comparison of the input impedances obtained with and without the on-chip grid inductance.**

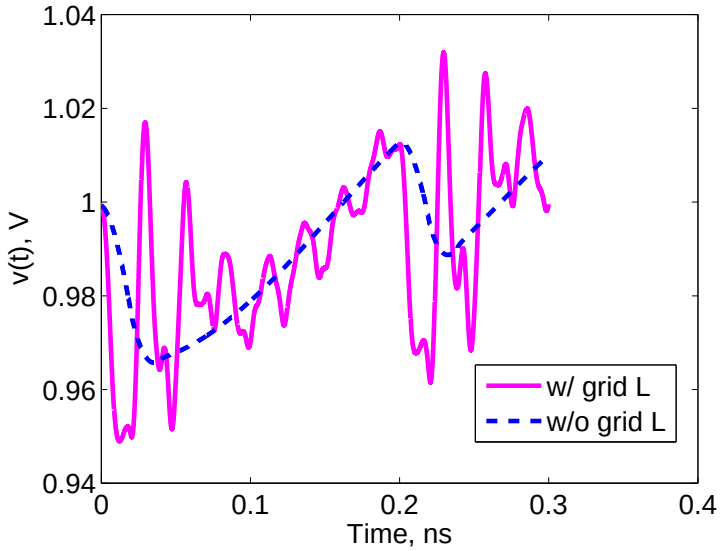



(a) Differential power-supply voltage at (x = 200 um, y = 200 um) in M1

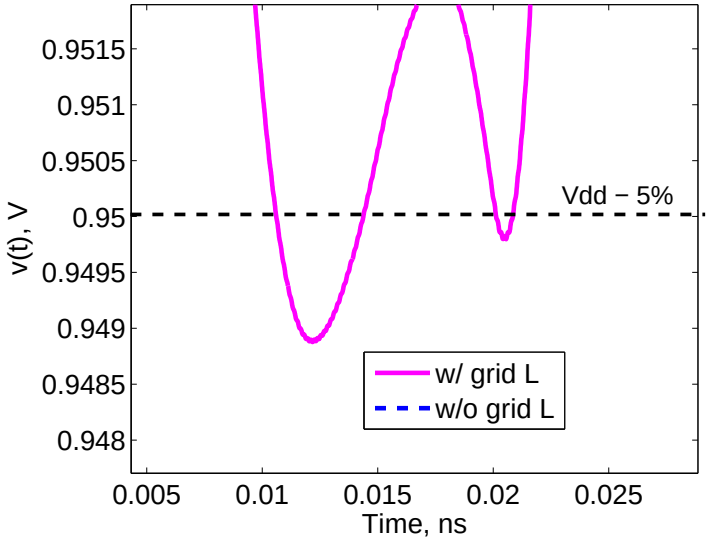



(b) Differential power-supply voltage at (x = 200 um, y = 60 um) in M1

**Figure 67. Comparison of power-supply voltage fluctuations obtained with and without the on-chip grid inductance.**




(a) Differential power-supply voltage at (x = 200 um, y = 180 um) in M1



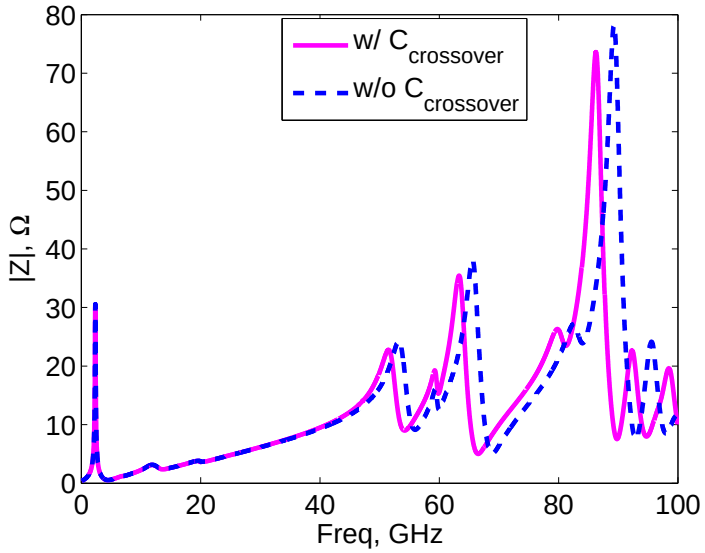

(b) Zoomed-in version of the differential power-supply voltage at (x = 200 um, y = 180 um) in M1

**Figure 68. Comparison of power-supply voltage fluctuations obtained with and without the on-chip grid inductance when a 80 pF decoupling capacitance is used.**

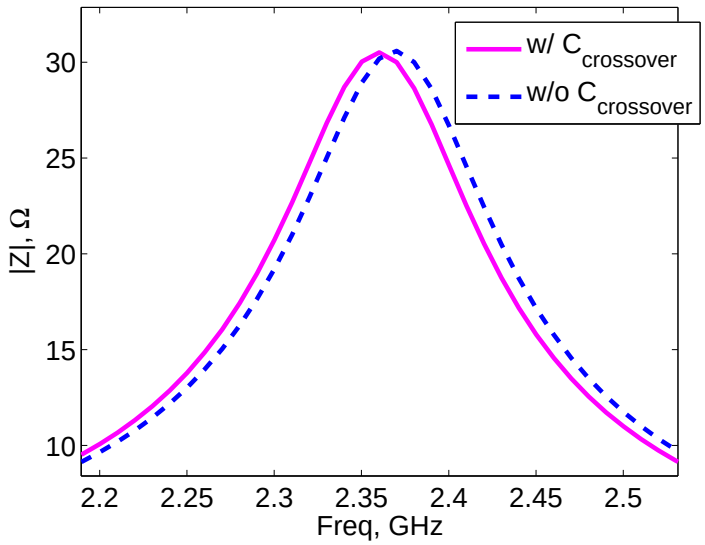
currents at these frequencies is far lesser in magnitude than the corresponding energy at small frequencies. From Figure 69(b), the decoupling nature of the crossover capacitance, though relatively small, can be observed. The crossover capacitance decreases the peak amplitude of the input impedance and lowers the resonant frequency, thereby serving as a decoupling capacitance. The effect of the crossover capacitance is not perceptible enough in the time domain. In Figure 70, the differential transient power-supply voltages are compared with and without the crossover capacitance. From Figure 70, it can be noticed that the crossover capacitance does not affect the PSN much.

## 7.6 Summary

In this chapter, LIM is extended to simulate PSN in on-chip power grids in the presence of on-chip decoupling capacitors. The need for a nonideal model for C4 bumps and the package PDN is demonstrated. A series RL model is employed to account for C4 bumps and the package PDN. The accuracy of LIM simulation is demonstrated in the presence of decoupling capacitors and nonideal model of C4 bumps and package PDN. Using the above changes to the equivalent circuit and the LIM-enabled transient simulation, the effect of the on-chip grid inductance on the PSN is studied. From the study, it is concluded that on-chip inductance does not affect the PSN significantly. However, it can induce sudden transient spikes in the power-supply voltage. Finally, the effect of the crossover capacitance on the PSN is also simulated. It has been observed that crossover capacitances act as decoupling capacitors, confirming the prior-held belief. However, because of their small values, it has been observed that they do not affect the PSN much.

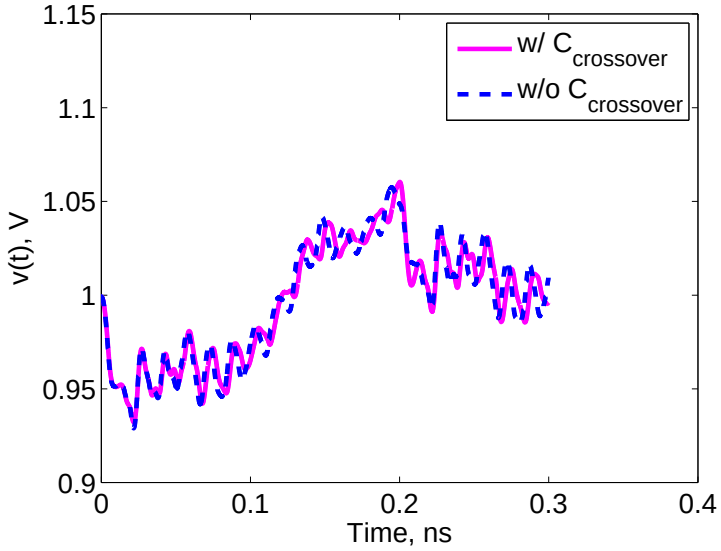



(a) Input impedance at the center of M1

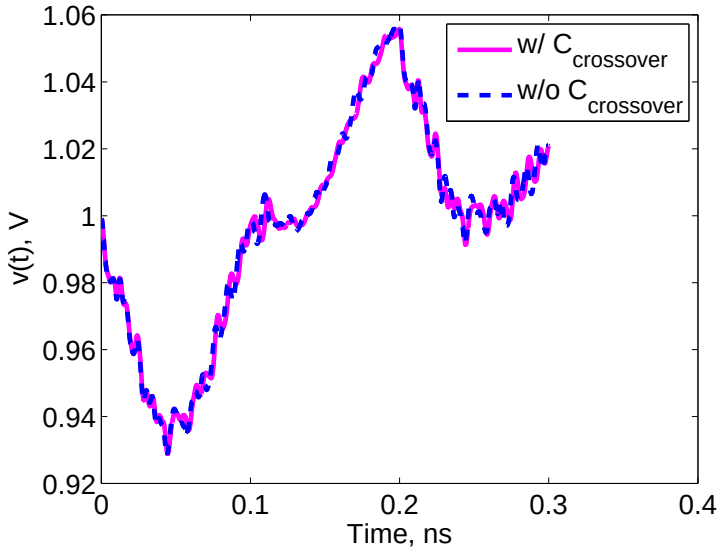



(b) Zoomed-in version of the input impedance near the chip-package resonance

**Figure 69. Comparison of the input impedances obtained with and without the crossover capacitance.**




(a) Differential power-supply voltage at (x = 200 um, y = 200 um) in M1




(b) Differential power-supply voltage at (x = 200 um, y = 60 um) in M1

**Figure 70. Comparison of power-supply voltage fluctuations obtained with and without the crossover capacitance.**

# CHAPTER 8

# ANALYTICAL STABILITY CONDITIONS OF THE LATENCY INSERTION METHOD FOR INHOMOGENEOUS GLC AND RLC CIRCUITS

## 8.1 Introduction

Until now, the working of the circuit-FDTD method and LIM for irregular on-chip PDNs has only been experimentally verified. However, this working has not yet been proven. This proof is the key to deriving/estimating the practical time complexity of LIM. For this proof, stability conditions of LIM for inhomogeneous circuits have to be proven. Until now, the stability conditions for only homogeneous RLC circuits have been proven [49]. However, the stability of any FDTD-like method for inhomogeneous circuits, be it circuit-FDTD method or LIM, is still an open problem (see [49]). It is desirable to get the stability conditions analytically so that the time step can be computed easily. *In this chapter, this open problem is partly solved: Analytical stability conditions of LIM are derived for the first time for inhomogeneous RLC (RLGC circuit without G)* [1] *and GLC (RLGC circuit without R) circuits.*[2] The conditions for RLC circuits are particularly important for on-chip PDN equivalent circuits. This derivation is made possible because of using Lyapunov's direct method (LDM), rather than the von Neumann method (VM), for stability analysis.

The rest of this chapter is organized as follows: In Section 8.2, the LIM formulation for the GLC circuits is described. In Section 8.3, the origin of the conditional stability in LIM is described. Also, in this section, the problem solved in this paper is defined mathematically. In Section 8.4, LDM is described. In Sections 8.5 and 8.6, analytical

[1] **S. N. Lalgudi**, M. Swaminathan, and Y. Kretchmer, "On-Chip Power Grid Simulation using Latency Insertion Method," Accepted for Publication in *IEEE Trans. on Circuits and Systems-I: Fundamental theory and applications*, 2008.

[2] **S. N. Lalgudi**, and M. Swaminathan, "Analytical Stability Condition of the Latency Insertion Method for Inhomogeneous GLC Circuits," Accepted for Publication in *IEEE Trans. on Circuits and Systems-II: Express Briefs*, 2008.

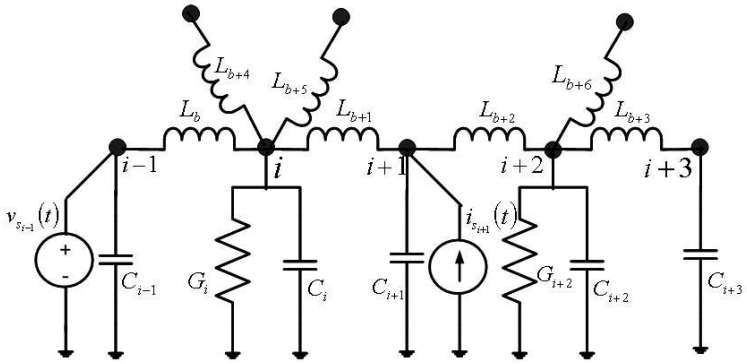



**Figure 71. An example of an inhomogeneous GLC circuit.**

stability conditions of LIM for inhomogeneous GLC and RLC circuits, respectively, are derived using LDM. Finally, in section 8.7, the conclusions of this chapter are drawn.

## 8.2 LIM-Based Transient Simulation Formulation for Inhomogeneous GLC Circuits

In this section, the GLC circuit, and the LIM formulation for the transient simulation of inhomogeneous GLC circuits are described.

A branch in a circuit is defined as a connection between two nodes excluding the ground reference node. To enable LIM in a circuit, 1) each branch in the circuit should have a nonzero inductance; otherwise, a small inductance is inserted into the branch to generate latency. 2) each node in the circuit should have a capacitance to ground; otherwise, a small shunt capacitance is added to generate latency at that node.

An example of an inhomogeneous GLC circuit is shown in Figure 71. Each inductor in this circuit is defined as a branch. Each node is marked as a solid black circle. The suffixes $i$ and $b$ denote a node and a branch, respectively. The quantity $L_b$ denotes the inductance of branch $b$; the quantities $C_i$ and $G_i$ denote the capacitance to ground and conductance to ground from node $i$, respectively. To enable LIM, $L_b > 0$ and $C_i > 0$. The quantity $G_i \geq 0$. Let $N_b^i$ denote the number of branches connected

to node $i$. In a homogeneous GLC circuit, there are some restrictions on the circuit elements's values and on the circuit topology: All branches should have the same inductance, i.e., $L_b = L$ for all $b$'s, and all nodes should have the same capacitance to ground and should have the same conductance to ground, i.e., $C_i = C$ and $G_i = G$ for all $i$'s. Moreover, each node is connected to same number of branches, i.e., $N_b^i = N$ for all $i$. In an inhomogeneous GLC circuit, there are no such restrictions (see Figure 71). Moreover, the quantity $N_b^i$ can be any positive integer. In Figure 71, the quantity $i_{s_i}(t)$ denotes a transient current source connected to node $i$, and $v_{s_i}(t)$ a transient voltage source connected to node $i$. The objective is to compute the transient node voltages computationally efficiently.

LIM is a transient simulation algorithm for circuits, similar to the FDTD method for dielectric media, and has optimal computational efficiency [48]. The LIM formulation for the transient simulation in GLC circuits (see Figure 71) is described next. Let $N_n$ denote the number of nodes and $N_b$ the number of branches. Let $\mathbf{C} \in \mathbb{R}^{N_n \times N_n}$ and $\mathbf{G} \in \mathbb{R}^{N_n \times N_n}$ denote the diagonal matrices of $C_i$'s and $G_i$'s, respectively. Let $\mathbf{L} \in \mathbb{R}^{N_b \times N_b}$ be the corresponding diagonal matrix of branch inductances. Let $v_i^{n+\frac{1}{2}}$ be the voltage of node $i$ at time instant $\left(n + \frac{1}{2}\right)\Delta t$, and let $\mathbf{v}^{n+\frac{1}{2}} \in \mathbb{R}^{N_n \times 1}$ be the vector of node voltages. Similarly, $i_b^n$ be the current in branch $b$ at time instant $n\Delta t$, and let $\mathbf{i}^n \in \mathbb{R}^{N_b \times 1}$ be the vector of branch currents. The LIM formulation involves obtaining update expressions for node voltages from the Kirchoff's current law (KCL) and obtaining update expressions for branch currents from the Kirchoff's voltage law (KVL) in an Yee-FDTD [46] manner.

The KCL at all nodes can be written as

$$\mathbf{C}\,\dot{\mathbf{v}}(t) + \mathbf{G}\mathbf{v}(t) = -\mathbf{M}^T\mathbf{i}(t) + \mathbf{i_s}(t), \tag{40}$$

where $\dot{v}(t) = \frac{dv(t)}{dt}$, the quantity $\mathbf{M}^T$ is the transpose of $\mathbf{M}$, and $\mathbf{M} \in \mathbb{Z}^{N_b \times N_n}$ is the edge-to-node incidence matrix. An entry in $\mathbf{M}$ corresponding to branch $b$ and node $i$

is defined as

$$\mathbf{M}(b,i) = \begin{cases} 1, & \text{if } i_b \text{ is flowing out of node } i \\ -1, & \text{if } i_b \text{ is flowing into node } i \\ 0, & \text{otherwise.} \end{cases}$$

When the KCLs in (40) are discretized using a semi-implicit integration scheme [49], the equation

$$\mathbf{C}\frac{\mathbf{v}^{n+\frac{1}{2}} - \mathbf{v}^{n-\frac{1}{2}}}{\Delta t} + \mathbf{G}\frac{\mathbf{v}^{n+\frac{1}{2}} + \mathbf{v}^{n-\frac{1}{2}}}{2} = -\mathbf{M}^T\mathbf{i}^n + \mathbf{i_s}^n \tag{41}$$

can be obtained. From (41), the node voltages are updated using the expression

$$\begin{aligned}(\mathbf{C} + 0.5\Delta t\ \mathbf{G})\,\mathbf{v}^{n+\frac{1}{2}} &= (\mathbf{C} - 0.5\Delta t\ \mathbf{G})\,\mathbf{v}^{n-\frac{1}{2}} - \Delta t\ \mathbf{M}^T\mathbf{i}^n \\ &\quad + \Delta t\ \mathbf{i_s}^n.\end{aligned} \tag{42}$$

Following a similar procedure, an update expression for the branch currents can be obtained. The KVLs in branches can be written as

$$\mathbf{L}\,\dot{\mathbf{i}}(t) = \mathbf{M}\mathbf{v}(t), \tag{43}$$

the discretized version of (43) can be written as

$$\mathbf{L}\frac{\mathbf{i}^{n+1} - \mathbf{i}^n}{\Delta t} = \mathbf{M}\mathbf{v}^{n+\frac{1}{2}}, \tag{44}$$

and the update expression for the branch currents can be obtained from (44) as

$$\mathbf{L}\mathbf{i}^{n+1} = \mathbf{L}\mathbf{i}^n + \Delta t\ \mathbf{M}\mathbf{v}^{n+\frac{1}{2}}. \tag{45}$$

The transient simulation using LIM involves computing the node voltages using (42) first and computing the branch currents using (45) next for each time step. When a node is connected to a voltage source, then as an intermediate step, the voltage of this node is made equal to the value of the voltage source at the current time instant.

The LIM transient simulation has optimal memory and time complexity (note the update process (42) and (45) involves only diagonal matrices) and has $\mathcal{O}\left((\Delta t)^2\right)$ accuracy (see [48]) However, this simulation is stable only for restricted values of $\Delta t$, i.e., the LIM formulation is only conditionally stable.

## 8.3 Conditional Stability and Stability Analysis of LIM

The discrete system described by (42) and (45) can be rewritten as

$$\mathbf{u}^{n+1} = \mathbf{A}^{-1}\mathbf{B}\mathbf{u}^{n} + \mathbf{A}^{-1}\mathbf{r}^{n}, \tag{46}$$

where the term $\mathbf{u}^n$ is the state vector, defined as

$$\mathbf{u}^{n} = \begin{bmatrix} (\mathbf{i}^{n})^{T} & \left(\mathbf{v}^{n-\frac{1}{2}}\right)^{T} \end{bmatrix}^{T},$$

the term $\mathbf{r}^n$ is the input vector, defined as

$$\mathbf{r}^{n} = \begin{bmatrix} (\mathbf{0})^{T} & \Delta t(\mathbf{i_s}^{n})^{T} \end{bmatrix}^{T},$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{L} & -\Delta t\ \mathbf{M} \\ \mathbf{0} & \mathbf{C} + 0.5\Delta t\ \mathbf{G} \end{bmatrix}, \tag{47}$$

and

$$\mathbf{B} = \begin{bmatrix} \mathbf{L} & \mathbf{0} \\ -\mathbf{\Delta t}\ \mathbf{M^T} & \mathbf{C} - 0.5\Delta t\ \mathbf{G} \end{bmatrix}. \tag{48}$$

The stability of the discrete system can be defined [81] 1) based on the boundedness of state, $\mathbf{u}^n$, given an initial condition for the state and a zero input, $\mathbf{r}^n$, 2) based on the bounded input bounded state (BIBS) stability, and 3) based on the bounded input bounded output (BIBO) stability. The focus of this paper is on the first kind of stability. This kind of stability is also a necessary condition for the BIBS stability [81], as it is a special case of the latter with $\mathbf{r}^n = \mathbf{0}$.

For the state stability, all the eigenvalues of the matrix $\mathbf{A}^{-1}\mathbf{B}$ should have a magnitude less than or equal to unity. In other words, the spectral radius of $\mathbf{A}^{-1}\mathbf{B}$, $\rho\left(A^{-1}B\right)$, is less than or equal to one. Unfortunately, the eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$ depend on $\Delta t$, resulting in a conditional stability of (46) dictated by the choice of $\Delta t$.

The conditions on $\Delta t$ can be computed by requiring $\rho\left(\mathbf{A}^{-1}\mathbf{B}\right) \leq 1$. However, finding eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$ analytically is a difficult problem. This difficulty is

avoided in some circuit toplogies if von Neumann method [82] is used for stability analysis [49]. In this method, the conditions on $\Delta t$ are determined by requiring the fourier amplitude of the state vector to be bounded by unity. The need to analyze the system in the Fourier domain requires the circuit element values to be equal and the circuit topology to be uniform at every point in the circuit. Specifically, analytical condition on $\Delta t$ is known and proven only for 1-D (i.e., $N_b^i = 2$) homogeneous RLC circuit (similar to homogeneous GLC circuit). Therefore, for an inhomogeneous GLC circuit, the stability condition cannot be derived using VM.

In [83], a similar problem in the FDTD method is solved for inhomogeneous lossy dielectric media. The approach in [83] is based LDM, introduced to the FDTD community in [84].

There are three important differences between the LIM problem and the FDTD problem [83] and [84]: 1) LIM discretizes only the circuits even when the circuits are discontinuous, while the FDTD problem discretizes both the dielectric medium and the free space. Therefore, the FDTD problem always solves a continuous problem domain. 2) Unlike the FDTD problem, the LIM problem can have more than three dimensions: in the LIM problem, the dimensions weakly refer to the number of branches connected to a node, which can be more than three. 3) Unlike the FDTD problem, the circuit problem is nonuniform with respect to the number of branches connected to a node.

The objective of this chapter is to obtain the conditions on $\Delta t$ for the state stability of LIM for inhomogeneous GLC and RLC circuits using the Lyapunov's direct method, discussed next for a discrete-time system.

## 8.4 Lyapunov's Direct Method (LDM) for Discrete-Time System

The stability of discrete-time systems can be analyzed using Lyapunov's direct method [81], which can be stated as follows: Let $\mathbf{u} \in \mathbb{R}^n$ be a vector of states of system, and $\mathbf{u} = \mathbf{0}$ be the equilibrium point. Suppose there exists a scalar function $E(\mathbf{u})$ continuous in $\mathbf{u}$ such that

$$E(\mathbf{0}) = 0, \quad E(\mathbf{u}) > 0 \ \text{ for } \ \mathbf{u} \neq \mathbf{0}, \ \text{ and} \tag{49}$$

$$E(\mathbf{u}(n\Delta t)) - E(\mathbf{u}(n-1)\Delta t) \leq 0 \ \text{ for all } \mathbf{u}. \tag{50}$$

Then, $\mathbf{u} = \mathbf{0}$ is stable. Moreover, if

$$E(\mathbf{u}(n\Delta t)) - E(\mathbf{u}(n-1)\Delta t) < 0 \ \text{ for } \ \mathbf{u} \neq \mathbf{0}, \tag{51}$$

then $\mathbf{u} = \mathbf{0}$ is asymptotically stable. If $E(\mathbf{u})$ satisfies (49) and (51) along with the condition that

$$||\mathbf{u}|| \longrightarrow \infty \quad \Longrightarrow \quad E(\mathbf{u}) \longrightarrow \infty, \tag{52}$$

then $\mathbf{u} = \mathbf{0}$ is globally asymptotically stable. The symbol $||\mathbf{u}||$ in (52) stands for the $p$-norm of the vector $\mathbf{u}$, where $p = 1, 2, \text{ and } \infty$. A continuous scalar function $E(\mathbf{u})$ satisfying (49) and (50) is called a Lyapunov function. Existence of a Lyapunov function is a sufficient condition for the stability of $\mathbf{u} = \mathbf{0}$.

## 8.5 Analytical Stability Condition for Inhomogeneous GLC Circuits

In this section, analytical stability condition for the LIM formulation in Section 8.2 is derived using LDM.

Since only the state stability is demonstrated, input (or excitations) are set to zero in (41). When $\mathbf{i_s}(t) = 0$, Equation (41) can be rewritten as

$$\mathbf{C}\frac{\mathbf{v}^{n+\frac{1}{2}} - \mathbf{v}^{n-\frac{1}{2}}}{\Delta t} = -\mathbf{M}^T\mathbf{i}^n - \mathbf{G}\frac{\mathbf{v}^{n+\frac{1}{2}} + \mathbf{v}^{n-\frac{1}{2}}}{2}. \tag{53}$$

For convenience, the discretized KVL equation (44) is repeated here:

$$\mathbf{L}\frac{\mathbf{i}^{n+1}-\mathbf{i}^{n}}{\Delta t}=\mathbf{M}\mathbf{v}^{n+\frac{1}{2}}. \tag{54}$$

The equilibrium state of the system (53)-(54) is same as that of (46). This state for (46) is the state for which $\mathbf{u}^{n+1}=\mathbf{u}^{n}$ for all $n$ in the absence of $\mathbf{r}^{n}$ (see [81], pp. 343). The origin $\mathbf{u_e}=\mathbf{0}$ is an equilibrium state of (46). In the following, an energy-like function is chosen as a scalar function, and the conditions for this function to be a Lyapunov function for the system (53)-(54) are determined. These conditions in turn result in an upper bound for $\Delta t$. When $\Delta t$ is chosen within this upper bound, $\mathbf{u_e}=0$ is stable.

A scalar function, $E^{n}$, is chosen as a potential candidate for the Lyapunov function for the system (53)-(54):

$$E^{n} \;=\; \tfrac{1}{2}\left(\mathbf{i}^{n-1}\right)^{T}\mathbf{L}\mathbf{i}^{n}+\tfrac{1}{2}\left(\mathbf{v}^{n-\frac{1}{2}}\right)^{T}\mathbf{C}\mathbf{v}^{n-\frac{1}{2}}. \tag{55}$$

The function $E^{n}$ can be shown to satisfy the condition in (50): Using (55), the quantity $E^{n}-E^{n-1}$ can be written as

$$\begin{aligned}
&E^{n}-E^{n-1}\\
&=\tfrac{1}{2}\left[\begin{array}{l}\left(\mathbf{i}^{n-1}\right)^{T}\mathbf{L}\mathbf{i}^{n}-\left(\mathbf{i}^{n-2}\right)^{T}\mathbf{L}\mathbf{i}^{n-1}+\\ \left(\mathbf{v}^{n-\frac{1}{2}}\right)^{T}\mathbf{C}\mathbf{v}^{n-\frac{1}{2}}-\left(\mathbf{v}^{n-\frac{3}{2}}\right)^{T}\mathbf{C}\mathbf{v}^{n-\frac{3}{2}}\end{array}\right]\\
&=\tfrac{1}{2}\left[\begin{array}{l}\left(\mathbf{i}^{n}+\mathbf{i}^{n-1}\right)^{T}\mathbf{L}\left(\mathbf{i}^{n}-\mathbf{i}^{n-2}\right)+\\ \left(\mathbf{v}^{n-\frac{1}{2}}+\mathbf{v}^{n-\frac{3}{2}}\right)^{T}\mathbf{C}\left(\mathbf{v}^{n-\frac{1}{2}}-\mathbf{v}^{n-\frac{3}{2}}\right)\end{array}\right],
\end{aligned}$$

which can be simplified using (53)-(54) as

$$\begin{aligned}
E^{n}-E^{n-1} \;&=\; -\tfrac{\Delta t}{4}\left(\mathbf{v}^{n-\frac{1}{2}}+\mathbf{v}^{n-\frac{3}{2}}\right)^{T}\mathbf{G}\left(\mathbf{v}^{n-\frac{1}{2}}+\mathbf{v}^{n-\frac{3}{2}}\right)\\
&\leq\; 0.
\end{aligned} \tag{56}$$

The inequality in (56) is true as $\mathbf{G}$ is positive semidefinite (semidefinite because conductances can be zero). Additionally, if $\mathbf{G}$ is positive definite, $E^{n}$ satisfies (51).

For $E^n$ to satisfy (49), $E^n$ is written as

$$\begin{aligned} E^n &= \tfrac{1}{2}\left(\mathbf{i}^{n-1}\right)^T \mathbf{L}\mathbf{i}^{n-1} + \tfrac{1}{2}\left(\mathbf{v}^{n-\frac{1}{2}}\right)^T \mathbf{C}\mathbf{v}^{n-\frac{1}{2}} + \\ &\quad \tfrac{\Delta t}{2}\left(\mathbf{i}^{n-1}\right)^T \mathbf{M}\mathbf{v}^{n-\frac{1}{2}} \\ &= \tfrac{1}{2}\begin{bmatrix} \mathbf{i}^{n-1} \\ \mathbf{v}^{n-\frac{1}{2}} \end{bmatrix}^T \begin{bmatrix} \mathbf{L} & \frac{\Delta t}{2}\mathbf{M} \\ \frac{\Delta t}{2}\mathbf{M^T} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{i}^{n-1} \\ \mathbf{v}^{n-\frac{1}{2}} \end{bmatrix} \\ &= \tfrac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x}. \end{aligned} \tag{57}$$

For $E_n$ to satisfy (49), the matrix $\mathbf{P}$ in (57) has to be positive definite. In the following, the conditions for $\mathbf{P}$ to be positive definite are found. The stability conditions are derived as a result. The stability conditions are found first when each node is connected to only two branches. These conditions are extended when the number of these branches is arbitrary and different for different nodes in the circuit.

### 8.5.1 Condition on $\Delta t$ when two branches are connected to every node

Let the subscript $i$ denote a node, and let the subscripts $i-\frac{1}{2}$ and $i+\frac{1}{2}$ denote the two branches connected to node $i$. Let $i^n_{i-\frac{1}{2}}$ and $i^n_{i+\frac{1}{2}}$ denote the branch currents that enter and leave node $i$, respectively. Let $b$ denote a branch, and let the two nodes of this branch be denoted by $(b,1)$ and $(b,2)$, with the branch current $i^n_b$ flowing from node $(b,1)$ to node $(b,2)$. The quantity $E^n$ in (57) can also be written as

$$\begin{aligned} E^n &= \tfrac{1}{2}\begin{bmatrix} \sum_{b=1}^{N_b} L_b\left(i_b^{n-1}\right)^2 + \sum_{i=1}^{N_n} C_i\left(v_i^{n-\frac{1}{2}}\right)^2 + \\ \Delta t \sum_{b=1}^{N_b} i_b^{n-1}\left(v_{b,1}^{n-\frac{1}{2}} - v_{b,2}^{n-\frac{1}{2}}\right) \end{bmatrix} \\ &= \sum_{b=1}^{N_b} \tfrac{1}{2}\begin{bmatrix} L_b\left(i_b^n\right)^2 + \frac{C_{b,1}}{2}\left(v_{b,1}^{n-\frac{1}{2}}\right)^2 + \frac{C_{b,2}}{2}\left(v_{b,2}^{n-\frac{1}{2}}\right)^2 + \\ \Delta t\ i_b^n\left(v_{b,1}^{n-\frac{1}{2}} - v_{b,2}^{n-\frac{1}{2}}\right) \end{bmatrix} \\ &= \sum_{b=1}^{N_b} E_b^n|_{(2,2)}, \end{aligned} \tag{58}$$

where

$$E_b^n|_{(p,q)} = \frac{1}{2}\begin{bmatrix} L_b\left(i_b^n\right)^2 + \frac{C_{b,1}}{p}\left(v_{b,1}^{n-\frac{1}{2}}\right)^2 + \frac{C_{b,2}}{q}\left(v_{b,2}^{n-\frac{1}{2}}\right)^2 + \\ \Delta t\ i_b^n\left(v_{b,1}^{n-\frac{1}{2}} - v_{b,2}^{n-\frac{1}{2}}\right) \end{bmatrix}. \tag{59}$$

From (58), $E^n$ is positive (in other words satisfies (49)) if $E_b^n|_{(2,2)}$ is positive for all $b$. Expressing $E_b^n|_{(2,2)}$ as a quadratic form

$$\begin{aligned} E_b^n|_{(2,2)} &= \frac{1}{2} \begin{bmatrix} i_b^{n-1} \\ v_{b,1}^{n-\frac{1}{2}} \\ v_{b,2}^{n-\frac{1}{2}} \end{bmatrix}^T \begin{bmatrix} L_b & \frac{\Delta t}{2} & -\frac{\Delta t}{2} \\ \frac{\Delta t}{2} & \frac{C_{b,1}}{2} & 0 \\ -\frac{\Delta t}{2} & 0 & \frac{C_{b,2}}{2} \end{bmatrix} \begin{bmatrix} i_b^{n-1} \\ v_{b,1}^{n-\frac{1}{2}} \\ v_{b,2}^{n-\frac{1}{2}} \end{bmatrix} \\ &= \frac{1}{2} \left(\mathbf{x}_b^n\right)^T \mathbf{P}_b \mathbf{x}_b^n. \end{aligned} \tag{60}$$

The quantity $E_b^n|_{(2,2)}$ is positive if the matrix $\mathbf{P}_b$ is positive definite. For $\mathbf{P}_b$ to be positive definite, all the upper left submatrices $\mathbf{P}_b^{(k)}$, where $k$ denotes the size of upper left submatrix, should have positive determinants [85]. The determinant of the first upper left matrix should then satisfy

$$\left|\mathbf{P}_b^{(1)}\right| = L_b > 0. \tag{61}$$

So all branch inductances should be nonzero and positive. Similarly, it can be shown that the condition $\left|\mathbf{P}_b^{(2)}\right| > 0$ is true if

$$(\Delta t)^2 < 4L_b \frac{C_{b,1}}{2}. \tag{62}$$

Since $(\Delta t)^2$ is non-negative for any real $\Delta t$, from (62), it can concluded that $C_i > 0$, i.e., all capacitances to ground should be positive. Finally, it can be shown that the condition $\left|\mathbf{P}_b^{(3)}\right| > 0$ is true if

$$(\Delta t)^2 < 4L_b \frac{\frac{C_{b,1}}{2}\frac{C_{b,2}}{2}}{\frac{C_{b,1}}{2} + \frac{C_{b,2}}{2}}. \tag{63}$$

Making use of the fact that

$$\frac{C_i C_{i+1}}{C_i + C_{i+1}} \geq \frac{1}{2} \min\left(C_i, C_{i+1}\right),$$

the condition in (63) is satisfied if

$$(\Delta t)^2 < 2L_b \min\left(\frac{C_{b,1}}{2}, \frac{C_{b,2}}{2}\right). \tag{64}$$

Since the condition in (64) is more strict than (62), the matrix $\mathbf{P}_b$ is positive definite if $L$'s and $C$'s are positive and (64) is satisfied. When this analysis is repeated for all $b$, the condition (64) becomes

$$\begin{aligned}(\Delta t)^2 &< 2\min_{b=1}^{N_b}\left(L_b \min\left(\tfrac{C_{b,1}}{2}, \tfrac{C_{b,2}}{2}\right)\right) \\ &< 2\min_{i=1}^{N_n}\left(\tfrac{C_i}{2}\min\left(L_{i+\frac{1}{2}}, L_{i-\frac{1}{2}}\right)\right),\end{aligned} \tag{65}$$

resulting in a condition for $\Delta t$ as

$$\Delta t < \sqrt{2}\min_{i=1}^{N_n}\left(\sqrt{\frac{C_i}{2}\min\left(L_{i-\frac{1}{2}}, L_{i+\frac{1}{2}}\right)}\right). \tag{66}$$

When (66) is simplified, the well-known Courant time step for 1-D circuit is obtained:

$$\Delta t < \min_{i=1}^{N_n}\left(\sqrt{C_i\min\left(L_{i-\frac{1}{2}}, L_{i+\frac{1}{2}}\right)}\right). \tag{67}$$

The proof for (52) for a positive function like $E^n$ is shown in Appendix B of [84]. Therefore, 1) when $\mathbf{G} \neq \mathbf{0}$ and $\Delta t$ satisfies (66), $\mathbf{u_e} = 0$ is globally asymptotically stable; 2) when $\mathbf{G} = \mathbf{0}$ and $\Delta t$ satisfies (66), $\mathbf{u_e} = 0$ is stable.

### 8.5.2 Condition on $\Delta t$ when arbitrary number of branches are connected to a node

Let $N_b^i$ denote the number of branches connected to node $i$. The generic condition on $\Delta t$ can be easily obtained by letting $p = N_b^{b,1}$ and $q = N_b^{b,2}$ in (59) and repeating the derivation from (58) through (66). For a generic case, the condition on $\Delta t$ in (66) can be shown to be

$$\Delta t < \sqrt{2}\min_{i=1}^{N_n}\left(\sqrt{\frac{C_i}{N_b^i}\min_{p=1}^{N_b^i}\left(L_{<i,p>}\right)}\right), \tag{68}$$

where $L_{<i,p>}$ denotes the value of $p$th inductor connected to node $i$. As can be observed, the derivation described thus far does not require the circuit to be homogeneous or infinitely long. Also, the (equivalent) circuits can be discontinuous, i.e., have irregularities in connections to neighboring nodes. Such discontinuities are observed in irregular on-chip power grids or in package power/ground planes with a hole.

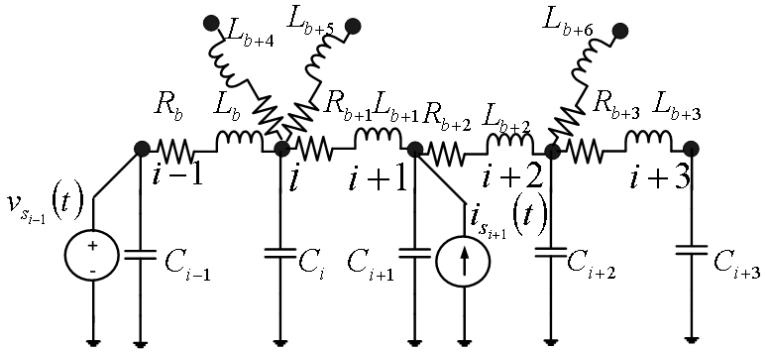



**Figure 72. An example of an inhomogeneous RLC circuit.**

## 8.6 Analytical Stability Condition for Inhomogeneous RLC Circuits

A RLC circuit (see Figure 72) is different from a GLC circuit (Figure 71) in two ways: there are no conductances to ground in the former, and there can be a resistance in series with the branch inductor. These differences in topology change the LIM formulation slightly: Due to the first difference, there will be no $\mathbf{G}$ term in (53). Due to the second, there will be a resistive $\mathbf{R}$ term in (54) whose form would be similar to the $\mathbf{G}$ term in (53). As a result, for a RLC circuit, the new node voltage update equation is

$$\mathbf{C}\frac{\mathbf{v}^{n+\frac{1}{2}}-\mathbf{v}^{n-\frac{1}{2}}}{\Delta t}=-\mathbf{M}^T\mathbf{i}^n, \tag{69}$$

and the new branch current update equation is

$$\mathbf{L}\frac{\mathbf{i}^{n+1}-\mathbf{i}^n}{\Delta t}=\mathbf{M}\mathbf{v}^{n+\frac{1}{2}}-\mathbf{R}\frac{\mathbf{i}^{n+1}+\mathbf{i}^n}{2}, \tag{70}$$

where $\mathbf{R}\in\mathbb{R}^{N_b\times N_b}$ is the diagonal matrix of branch resistances.

The new update expressions (69) and (70) too impose a conditional stability on the results. The conditions on $\Delta t$ for a RLC circuit is shown below to be same as the one in (68).

The Lyapunov function, $E^n$ (see 55), for GLC circuits is not a Lyapunov for RLC circuits. As a result, a new scalar function, $F^n$, is proposed:

$$F^n=\frac{1}{2}\left(\mathbf{i}^n\right)^T\mathbf{L}\mathbf{i}^n+\frac{1}{2}\left(\mathbf{v}^{n-\frac{1}{2}}\right)^T\mathbf{C}\mathbf{v}^{n+\frac{1}{2}}. \tag{71}$$

The function $F^n$ in (71) can be shown to satisfy condition in (50) for the system (69)-(70): Using (71), the difference in $F^n$ between successive time instants can be written as

$$\begin{aligned} & F^n - F^{n-1} \\ &= \frac{1}{2}\begin{bmatrix} (\mathbf{i}^n)^T \mathbf{L}\mathbf{i}^n - (\mathbf{i}^{n-1})^T \mathbf{L}\mathbf{i}^{n-1} + \\ \left(\mathbf{v}^{n-\frac{1}{2}}\right)^T \mathbf{C}\mathbf{v}^{n+\frac{1}{2}} - \left(\mathbf{v}^{n-\frac{3}{2}}\right)^T \mathbf{C}\mathbf{v}^{n-\frac{1}{2}} \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} (\mathbf{i}^n + \mathbf{i}^{n-1})^T \mathbf{L}\,(\mathbf{i}^n - \mathbf{i}^{n-1}) + \\ \left(\mathbf{v}^{n-\frac{1}{2}}\right)^T \mathbf{C}\left(\mathbf{v}^{n+\frac{1}{2}} - \mathbf{v}^{n-\frac{3}{2}}\right) \end{bmatrix}, \end{aligned}$$

which can be simplified using (69) and (70) as

$$\begin{aligned} F^n - F^{n-1} &= -\tfrac{\Delta t}{4}(\mathbf{i}^n + \mathbf{i}^{n-1})^T \mathbf{R}\,(\mathbf{i}^n + \mathbf{i}^{n-1}) \\ &\leq 0. \end{aligned} \tag{72}$$

The inequality in (72) is true as $\mathbf{R}$ is positive semidefinite (note resistances can be zero). Additionally, if $\mathbf{R}$ is positive definite, $F^n$ satisfies (51).

Like in Section 8.5, the function $F^n$ can be written as

$$\begin{aligned} F^n &= \frac{1}{2}\begin{bmatrix} \mathbf{i}^n \\ \mathbf{v}^{n-\frac{1}{2}} \end{bmatrix}^T \begin{bmatrix} \mathbf{L} & -\frac{\Delta t}{2}\mathbf{M} \\ -\frac{\Delta t}{2}\mathbf{M}^T & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{i}^n \\ \mathbf{v}^{n-\frac{1}{2}} \end{bmatrix} \\ &= \tfrac{1}{2}\mathbf{y}^T\mathbf{Q}\mathbf{y}. \end{aligned} \tag{73}$$

It can be noticed that the matrix $\mathbf{Q}$ in (73) is different from the matrix $\mathbf{P}$ in (57). Just like in the previous section, the function $F^n$ can be shown to satisfy (49) if $\mathbf{Q}$ is shown to be positive definite. Though matrix $\mathbf{Q}$ is different from matrix $\mathbf{P}$, it is shown below that the conditions for $\mathbf{Q}$ to be positive definite are same as those for $\mathbf{P}$.

### 8.6.1 Condition on $\Delta t$ when two branches are connected to every node

When each node is connected to only two branches, the quantity $F^n$ in (73) can be rewritten as

$$\begin{aligned} F^n &= \frac{1}{2}\begin{bmatrix} \sum_{b=1}^{N_b} L_b \left(i_b^n\right)^2 + \sum_{i=1}^{N_n} C_i \left(v_i^{n-\frac{1}{2}}\right)^2 + \\ \Delta t \sum_{i=1}^{N_n} v_i^{n-\frac{1}{2}} \left(i_{i-\frac{1}{2}}^n - i_{i+\frac{1}{2}}^n\right) \end{bmatrix} \\ &= \sum_{b=1}^{N_b} \frac{1}{2}\begin{bmatrix} L_b \left(i_b^n\right)^2 + \frac{C_{b,1}}{2}\left(v_{b,1}^{n-\frac{1}{2}}\right)^2 + \frac{C_{b,2}}{2}\left(v_{b,2}^{n-\frac{1}{2}}\right)^2 - \\ \Delta t\ i_b^n \left(v_{b,1}^{n-\frac{1}{2}} - v_{b,2}^{n-\frac{1}{2}}\right) \end{bmatrix} \\ &= \sum_{b=1}^{N_b} F_b^n|_{(2,2)}, \end{aligned} \tag{74}$$

where

$$F_b^n|_{(p,q)} = \frac{1}{2}\begin{bmatrix} L_b \left(i_b^n\right)^2 + \frac{C_{b,1}}{p}\left(v_{b,1}^{n-\frac{1}{2}}\right)^2 + \frac{C_{b,2}}{q}\left(v_{b,2}^{n-\frac{1}{2}}\right)^2 - \\ \Delta t\ i_b^n \left(v_{b,1}^{n-\frac{1}{2}} - v_{b,2}^{n-\frac{1}{2}}\right) \end{bmatrix}. \tag{75}$$

It can be noticed that $F_b^n|_{(p,q)}$ in (75) is different from $E_b^n|_{(p,q)}$ in (59), in that the former has a negative sign, instead of a positive sign, before the third term in (75).

Since $F_b^n|_{(p,q)}$ is different from $E_b^n|_{(p,q)}$, the conditions for $F_b^n|_{(p,q)}$ to be positive should be shown to be same as those for $E_b^n|_{(p,q)}$. Pursuant to this aim, the function $F_b^n|_{(2,2)}$ is expressed as a quadratic form similar to the way $E_b^n|_{(2,2)}$ was expressed:

$$\begin{aligned} F_b^n|_{(2,2)} &= \frac{1}{2}\begin{bmatrix} i_b^n \\ v_{b,1}^{n-\frac{1}{2}} \\ v_{b,2}^{n-\frac{1}{2}} \end{bmatrix}^T \begin{bmatrix} L_b & -\frac{\Delta t}{2} & \frac{\Delta t}{2} \\ -\frac{\Delta t}{2} & \frac{C_{b,1}}{2} & 0 \\ \frac{\Delta t}{2} & 0 & \frac{C_{b,2}}{2} \end{bmatrix} \begin{bmatrix} i_b^n \\ v_{b,1}^{n-\frac{1}{2}} \\ v_{b,2}^{n-\frac{1}{2}} \end{bmatrix} \\ &= \frac{1}{2}\left(\mathbf{y}_b^n\right)^T \mathbf{Q}_b \mathbf{y}_b^n. \end{aligned} \tag{76}$$

It can be noticed that the matrix $\mathbf{Q}_b$ in the above equation is different from $\mathbf{P}_b$ in Section 8.5. The quantity $F_b^n|_{(2,2)}$ is positive if the matrix $\mathbf{Q}_b$ is positive definite. It can be verified that the determinants of $\mathbf{Q}_b^{(k)}$, $k = 1, 2, 3$, are same as the corresponding determinants of $\mathbf{P}_b^{(k)}$. Therefore, the condition for $\mathbf{Q}_b$ to be positive definite is same those for $\mathbf{P}_b$, which is given in (67). Therefore, if $\Delta t$ satisfies (67), the function $F_b^n|_{(p,q)}$ is positive. As a result, the scalar function $F^n$ in (74) is also positive.

Therefore, $F^n$ satisfies (49) too and is therefore a Lyapunov function for the system (69)-(70).

The proof for (52) for a positive function like $F^n$ is same as that for $E^n$. Therefore, 1) when $\mathbf{R} \neq \mathbf{0}$ and $\Delta t$ satisfies (66), the origin is globally asymptotically stable; 2) when $\bar{\bar{\mathbf{R}}} = \bar{\bar{\mathbf{0}}}$ and $\Delta t$ satisfies (66), the origin is stable.

The condition on $\Delta t$ when more than two branches are connected to a node is same as that in (68). This condition can be shown with $F^n$ following the same procedure as in Section 8.5 with $E^n$.

## 8.7 Summary

Latency insertion method (LIM) is a transient simulation technique for circuits and is based on a finite-difference formulation, like the well-known finite-difference time-domain (FDTD) method for solving Maxwell's equations. LIM, like the FDTD method, is only conditionally stable, resulting in an upper bound for the time step of the transient simulation. This bound on the time step is a function of the circuit topology and circuit element values. It is critical to know this bound analytically for a given circuit. Stability conditions of the LIM have been proven only for inhomogeneous RLC circuits. In this paper, analytical stability conditions of LIM for inhomogeneous GLC and RLC circuit are derived for the *first time*, resulting in an upper bound for the time step. This upper bound for time step has been observed experimentally.

# CHAPTER 9

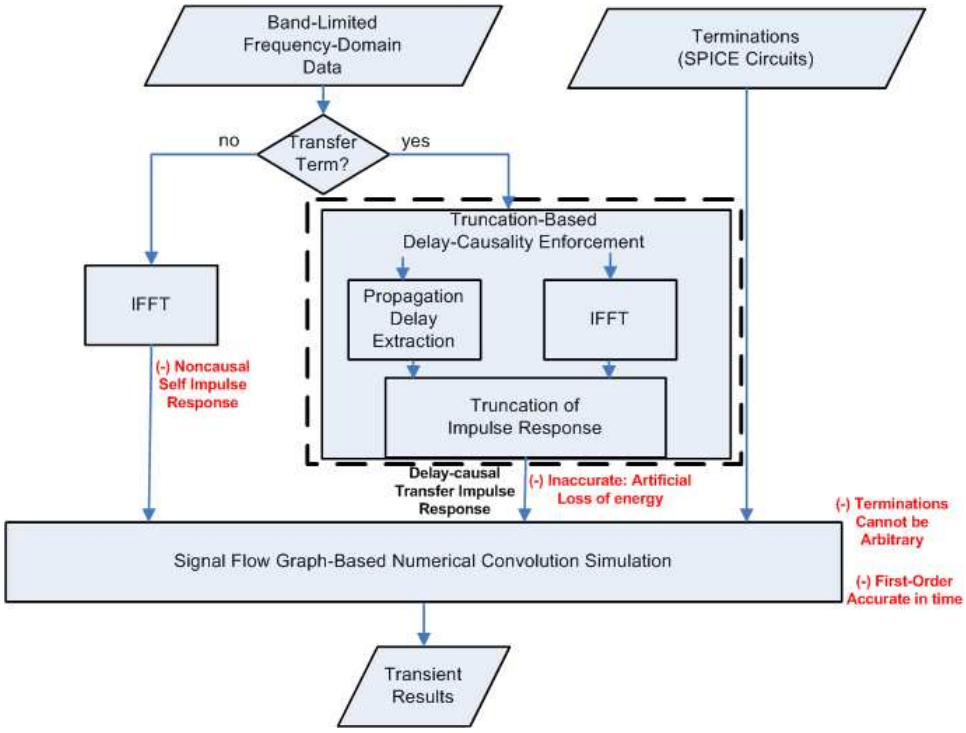
# DELAY CAUSALITY ENFORCEMENT THROUGH NONMINIMUM-PHASE RECONSTRUCTION-BASED TECHNIQUE

## 9.1 Introduction

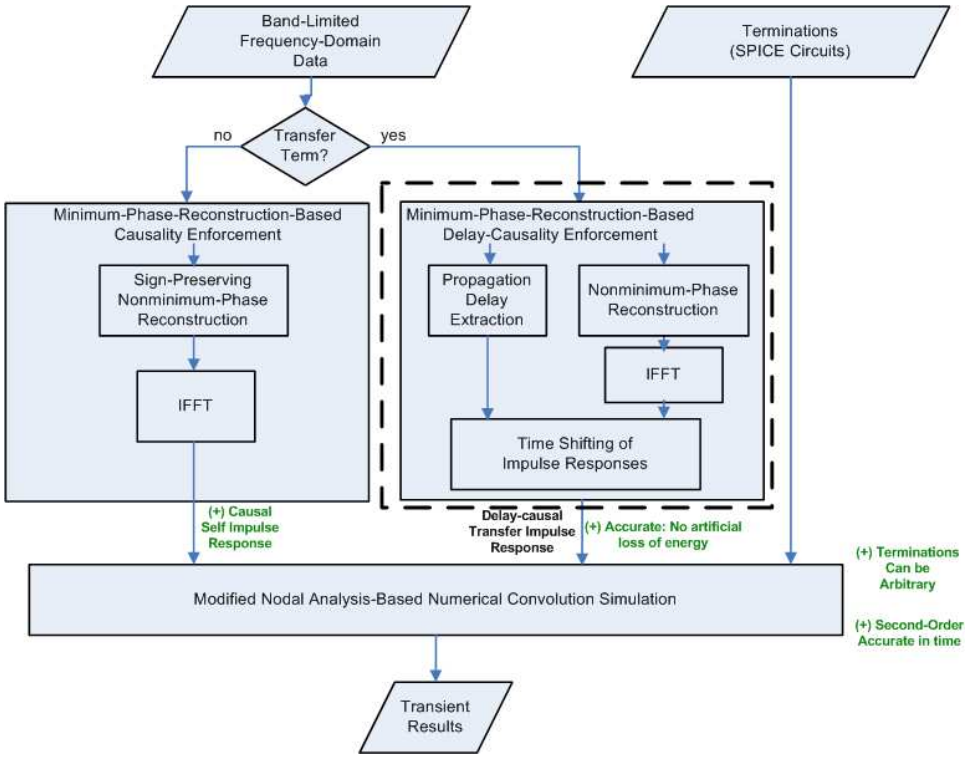
This chapter is the beginning of the second part of this dissertation. This part is about developing a time-domain technique for the causal transient simulation of b.l.f.d. data. In Chapter 1, the importance of enforcing delay-causality in transfer frequency responses was described. The approach proposed for this purpose in the prior work [55], [56], [57] degrades the accuracy of transient results. In this chapter, a new approach for this purpose is proposed. This approach has been demonstrated not to suffer from the inaccuracy limitations of the approach in [55], [56], [57]. The focus of this chapter is also shown in Figure 73. An improved version of the proposed approach has been described in Chapter 10.

## 9.2 Brief Background

Numerical convolution-based approach is common for the transient simulation of interconnects characterized by multiport b.l.f.d. data. To capture the propagation delay between ports, the port-to-port impulse response has to be zero for times less than the port-to-port propagation delay. Such an impulse response is said to be delay-causal. When the frequency-domain data are known only for a finite bandwidth, the port-to-port impulse responses are usually not delay-causal. Two techniques are common to obtain delay-causal impulse response from band-limited data. In the first technique, the part of the impulse response before the delay is zeroed or truncated. The first technique is the popular choice to enforce delay causality [55], [57]. In the




(a) Prior approach [55], [56], [57].




(b) Proposed approach in this chapter. Note that the delay-causality enforcement technique shown in this figure is different from the proposed approach by a sign term.

**Figure 73. Comparison of the prior and proposed approach in numerical-convolution-based causal transient simulation of band-limited data. The focus of this chapter is the region marked within the dashed rectangle.**

second technique, a causal impulse response is first obtained through a minimum-phase reconstruction of the data and is shifted in time to account for the propagation delay. As only modified responses are produced, accuracy is affected in both the techniques. The transient results from these techniques are considered comparable, and sometimes, equivalent.

In this chapter, the accuracy of the transient results from these two techniques is compared. It has been shown that the transient solutions from the techniques are not equivalent in most cases. Specifically, *it has been shown that the truncation-based technique [55], [57] does not preserve the energy in port-to-port frequency responses and the accuracy of the simulation can be poor, while the minimum-phase reconstruction-based technique does not have this inaccuracy problem.* Also, it has been shown that (not only the band-limited nature of the data but also) the frequency-domain windowing can make the impulse response nondelay-causal. Minimum-phase reconstruction-based technique has been applied with reasonable success to handle even delay-causality violations because of noncausal data and frequency-domain windowing.

## 9.3 Background

For many interconnects, the frequency-domain data (e.g., scattering parameters) are known for a limited bandwidth. The objective is to perform a transient simulation of such (multiport) data along with the (port) terminations. Propagation delay through the interconnects may not be accurately captured when the data are noncausal [59] or when the data are band limited [55], [56], [86]. In [55], [56], [57], [86], numerical convolution-based approaches that capture the delay from a band-limited data have been proposed.

Port-to-port propagation delays are captured accurately, if the port-to-port time-domain impulse responses are (made) delay-causal [55], [57], [86]. In [57], [86], a

causal impulse response is first constructed from the minimum-phase component of the data (which requires the use of the Hilbert transform on the magnitude of the data [87]), the delay is extracted subsequently, and the impulse response is shifted by this delay, producing a delay-causal response. The transient response from this technique, however, may not be the actual transient response of the system [87]. The accuracy of the simulation using this technique is not clearly understood.

In [55], a truncation-based technique has been proposed, which avoids the inaccuracy problem associated with [57], [86]. In this technique, the delay is extracted as it was in [57], [86]; the nondelay-causal impulse responses are obtained by inverse fourier transform (IFT) of the data; and the nondelay-causal part of the responses are truncated (or zeroed), producing a delay-causal response. Since the data are not reconstructed in this technique, the above inaccuracy problem is avoided. However, since the impulse responses are modified with time-domain truncation, the accuracy of simulation may still be affected. However, in [57], the results from both the techniques are treated equivalent. Therefore, understanding the effects of this technique on accuracy becomes important. Moreover, the performance of both the techniques is not clearly understood in the presence of other sources of delay-causality violations (e.g., noncausal data). This understanding is important when the causality of the data is not known.

In this chapter, a comparative study of the accuracy and the performance of the transient simulation using the two techniques has been carried out in the presence of delay-causality violations. Towards to this end, frequency-domain windowing has been shown as a new source of delay-causality violation. Prior work in windowing has been on its need and on the choice of functions that does not affect the accuracy of transient simulation much [88]. From the study, the following have been found:

1. When delay-causality is not enforced, transient results can be inaccurate. When

the data are causal, this inaccuracy can be improved by increasing the bandwidth of the data. On the other hand, when the data are not causal, such an improvement may not be possible.

2. Truncation-based technique does not preserve the energy in the original frequency response, and the transient results from this technique can be inaccurate. This inaccuracy can be alleviated by increasing the bandwidth of the data only if the data are already causal. This inaccuracy can be significant when the data are noncausal. The accuracy can degrade further when frequency-domain windowing is also applied.

3. Minimum-phase-based technique preserves the energy in the original frequency response, and the transient results from this technique can be reasonably accurate even when the data are band limited, are noncausal, and when frequency-domain windowing is applied. For an ideal transmission line, the results are exact.

The rest of this chapter is organized as follows: In Section 9.4, the different sources of delay-causality violations and the difference in accuracy between the techniques are described. In Section 9.5, numerical results demonstrating the performance of the techniques are shown. Finally, in Section 9.6, the conclusions of this chapter are presented.

## 9.4 Delay-Causality Violations and Delay-Causality Enforcement

Consider a linear time-invariant system, with propagation delay $t_p$, whose frequency response, $H(f)$, is known at uniformly-spaced frequencies between zero and a maximum frequency, $f_c$. Let $h(t)$ be the IFT of $H(f)$. The response $h(t)$ is delay-causal for a physical system. Let $\breve{H}(f)$ denote $H(f)$ when the latter is not causal. Let $G(f)$

and $W(f)$ denote gate function and window function [88], respectively, with a cut-off frequency $f_c$. Then the frequency response before IFT, $\hat{H}(f)$, can be expressed as

$$\hat{H}(f) = \breve{H}(f)\, G(f)\, W(f). \tag{77}$$

Let $\breve{h}(t)$, $g(t)$, and $w(t)$ be the the IFTs of the $\breve{H}(f)$, $G(f)$, and $W(f)$, respectively. Then from (77), $\hat{h}(t)$ can be written as

$$\hat{h}(t) = \breve{h}(t) * g(t) * w(t), \tag{78}$$

where the symbol $'*'$ denotes the linear convolution operator. Both $G(f)$ and $W(f)$ are real and even functions in $f$, therefore, $g(t)$ (a sinc function) and $w(t)$ are real and even functions in $t$, making them noncausal. They become an impulse function, $\delta(t)$, and thereby causal, only for $f_c = \infty$. Also, $\breve{h}(t)$ is not delay-causal. As a result, $\hat{h}(t)$ is usually not delay-causal. The transient results from a nondelay-causal $\hat{h}(t)$ can be inaccurate (see Figure 76). The band-limited delay-causality violations [55], [56], [57], [86] are because of $G(f)$. Unlike $G(f)$, $W(f)$ has two unwanted effects: 1) It can make $\hat{h}(t)$ more nondelay-causal (see Figure 83); 2) In some cases, it can make a delay-causal $\hat{h}(t)$ nondelay-causal: If $H(f)$ is the transfer S-parameter of an ideal transmission line with $f_c = \frac{k}{t_p}$, $\Delta t = 1/(2f_c)$, $t_p = m\Delta t$, where $m$ and $k$ are positive integers, and $\Delta t$ is the time step of transient simulation. In such a case, $\hat{h}(t) = \delta(t - t_p)$ without $W(f)$. However, with $W(f)$, $\hat{h}(t) = w(t - t_p)$, a nondelay-causal response. For small $t_p$, $w(t - t_p)$ can become noncausal too. The delay-causality violations because of $G(f)$ and $W(f)$ can be alleviated by increasing $f_c$ if $\breve{H}(f) = H(f)$ (see Figure 77) is ensured first. Such an alleviation is not possible if $\breve{H}(f) \neq H(f)$ (see Figure 81). Hence, obtaining a delay-causal impulse response, $\tilde{h}(t)$, from $\hat{H}(f)$ is essential. In [55], [56], [57], [86], the violations treated are mostly because of $G(f)$.

### 9.4.1 Delay Extraction From Band-Limited Data

In [55], [56], [57], a numerical procedure to extract the propagation delay from the b.l.f.d. data has been proposed. This procedure is described in this subsection. In [55],

[56], a decomposition procedure for $H(\omega)$ has been presented. This decomposition is given as

$$H(\omega) = H_{\min}(\omega)\, e^{-j\omega t_p}, \tag{79}$$

where $H_{\min}(\omega)$ is the minimum-phase component (see [89], [87]). The function $H_{\min}(\omega)$ is the delayless part of $H(\omega)$ and models effects because of attenuation and dispersion. The term $e^{-j\omega t_p}$ is the all-pass component, which has been used to model the phase using a linear-phase condition. The term $t_p$ here refers to the propagation delay if $H(\omega)$ was lossless. For a lossy transmission line of length $l$, $t_p$ would mean the propagation delay of a lossless line of the same length and is calculated by the value of the propagation delay at $\omega = \infty$ [90].

The component $H_{\min}(\omega)$ in (86) is computed from the magnitude of $H(\omega)$:

$$|H_{\min}(\omega)| = |H(\omega)|; \tag{80}$$

$$\arg[H_{\min}(\omega)] = -\mathrm{HT}\{\ln|H(\omega)|\}. \tag{81}$$

In (80) and (81), $|x|$ stands for the magnitude of $x$, $\arg[x]$ is the principal argument of complex number $x$, and $\mathrm{HT}\{\mathrm{x}\}$ is the Hilbert transform [89] of $x$. Using a discrete Hilbert transform [89], (89) can be rewritten as

$$\arg[H_{\min}(\omega)] = -\frac{1}{2\pi}\mathcal{P}\int_{\theta=-\pi}^{\pi} \ln|H(\theta)| \cot\left(\frac{\omega-\theta}{2}\right) d\theta, \tag{82}$$

where $\mathcal{P}$ denotes the Cauchy principal value of the integral that follows. In [55], [56], the delay $t_p$ in (79) is computed as

$$t_p = -\mathrm{AverageSlope}\left(\arg\left[\frac{\mathrm{H}(\omega)}{\mathrm{H}_{\min}(\omega)}\right]\right). \tag{83}$$

Propagation delay computed in the average sense (83) works well when there is just a single delay in the frequency response. When more than one delay is present in the frequency response, like in a dispersive transmission line where the velocity is a function of frequency, then $t_p$ calculated from (83) is an approximation. For a more accurate $t_p$, $t_p$ should be computed as the smallest delay in the response.

### 9.4.2 Truncation-based Delay-Causality Enforcement

In [55], [56], [57], delay-causal impulse responses are obtained through a truncation-based technique, which is described in this section. In a truncation-based technique, $\tilde{h}(t)$ is obtained from $\hat{h}(t)$ by forcing $\hat{h}(t)$ to be zero for $t < t_p$:

$$\tilde{h}(t) = \hat{h}(t)\,\phi(t), \tag{84}$$

where $\phi(t)$ is 1 for $t \geq t_p$ and 0 for $t < t_p$. Because of $\phi(t)$, $\tilde{h}(t) \neq \hat{h}(t)$ usually (see Figure 75(a)), and their fourier transforms (a measure of energy) may be different at all frequencies (see Figure 75(b)). One noticeable difference could be the DC values ($\propto$ fourier transform value at $f = 0$) of the transient responses (see Figure 79(a) or Figure 82). The value of the fourier transform at $f = 0$ is equal to the area under the time-domain curve (this follows from the definition of Fourier transform). Because of $\phi(t)$, the area under $\hat{h}(t)$ may not be the same as that under $\tilde{h}(t)$. The areas would be the same if $\hat{h}(t)$ is delay-causal. However, when $\hat{h}(t)$ is nondelay-causal (which is often the case), the areas will differ. This disparity represents a nonpreservation of energy. This disparity can be alleviated by increasing $f_c$ only if $\hat{h}(t)$ becomes (more) delay-causal in the process. When $\breve{H}(f) \neq H(f)$, however, this disparity can be significant (see Figure 82) and can worsen when windowing is also applied (see Figure 84).

### 9.4.3 Minimum-Phase Reconstruction-based Delay-Causality Enforcement

From (79), a delay-causal impulse response can also be obtained. The IFT of the b.l. response $H_{\min}(0 : \omega_c)$, $\hat{h}_{\min}(t)$, is causal (see [89]). Therefore, the IFT of the R.H.S. of (86), $\hat{h}_{\min}(t - t_p)$, is delay-causal. Therefore, $\tilde{h}(t)$ is chosen as

$$\tilde{h}(t) = \hat{h}_{\min}(t - t_p). \tag{85}$$

This procedure to enforce delay-causality is referred to as the minimum-phase reconstruction-based delay-causality enforcement. This chapter proposes this procedure for the delay-causality enforcement over the truncation-based procedure proposed in Section

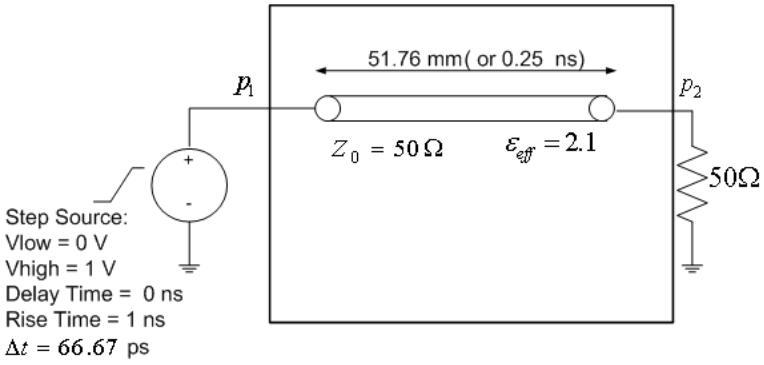



**Figure 74. Test setup of a step response of a lossless transmission line, $t_p = 0.25$ ns.**

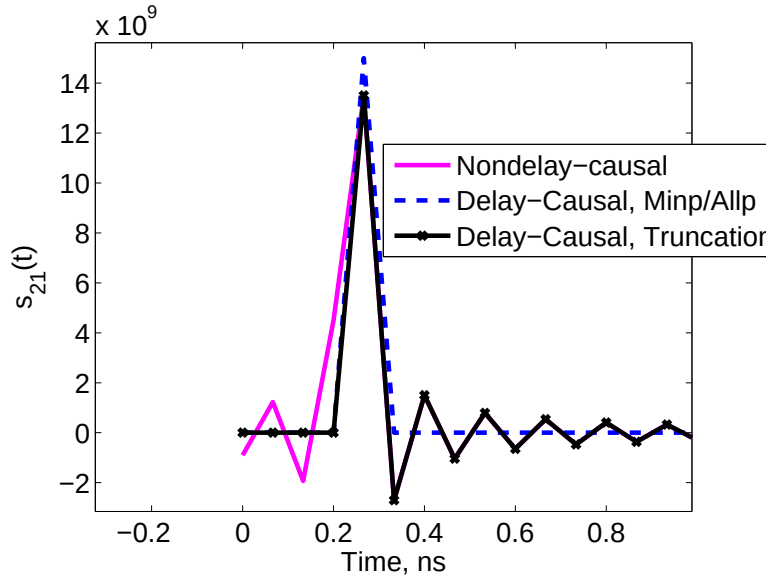
9.4.2. Unlike the truncation-based method (Section 9.4.2), a $\tilde{h}(t)$ can be produced with the same energy in $\hat{H}(f)$ even when $\breve{H}(f) \neq H(f)$ (see Figure 82) and $W(f)$ is present (see Figure 85).

## 9.5 Results

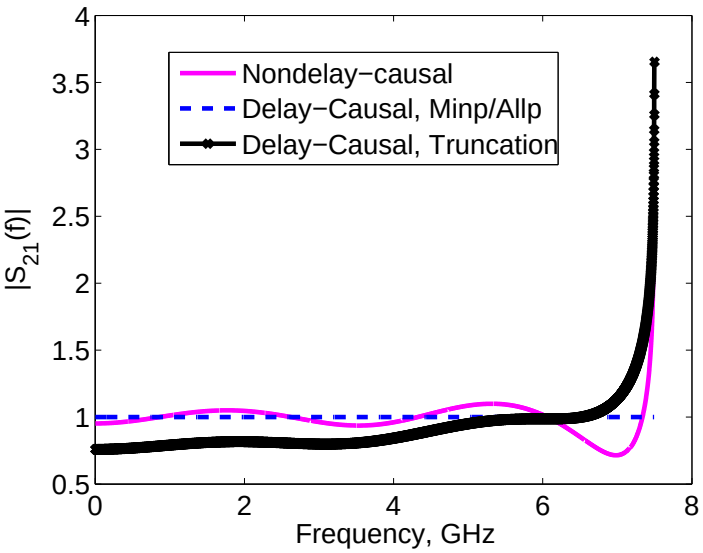
In this section, the accuracy of the transient simulation using the two techniques are compared in the presence of delay-causality violations because of $\breve{H}(f)$, $G(f)$, and $W(f)$. In the examples below, the step responses of transmission lines are simulated. The far-end voltages are computed. The results from Agilent ADS ('ADS' in Figure 76 and Figure 82) are used for comparison. The following cases are considered:

1. $\breve{H}(f) = H(f)$ and no $W(f)$ (Causal data, see Figures 74-79):

   $\breve{H}(f) = H(f)$ is ensured with an ideal transmission line ($S_{11}(f) = 0$, $S_{21}(f) = \exp(-j2\pi f t_p)$ for $f \leq f_c$), with $t_p = 0.25$ ns (see Figure 74). In Figure 75, the results with $\hat{h}(t)$ and $\tilde{h}(t)$ are compared for $f_c = 7.5$ GHz. From Figure 75(a), it can be seen that nondelay-causal $\hat{h}(t)$ is also not causal (see the nonzero value at $t = 0$). A noncausal $\hat{h}(t)$ is inherently truncated by causal convolution integral, $\int_{\tau=0}^{t} \hat{h}(\tau) x(t-\tau)\, d\tau$, and hence the transient results from such a $\hat{h}(t)$ can be inaccurate (see Figure 76(a-b)). Because of truncation, the fourier transforms of $\hat{h}(t)$ and truncation-based $\tilde{h}(t)$ are different from that of minimum-phase-based

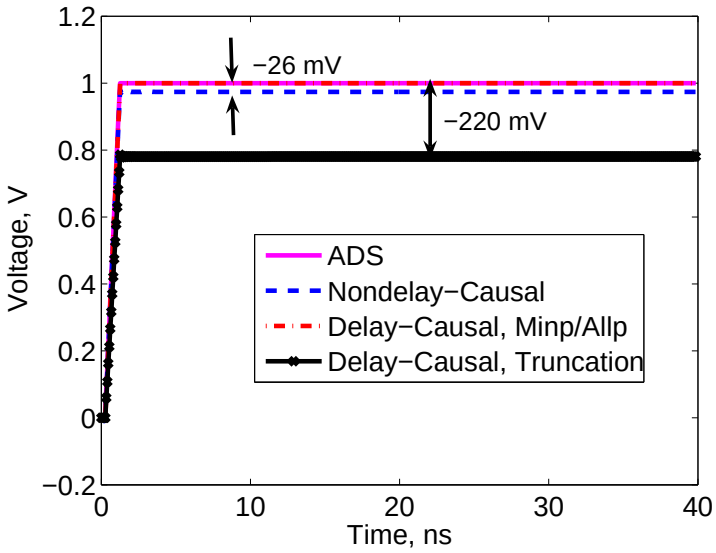



(a) Transfer impulse responses, $t_p = 0.25$ ns.

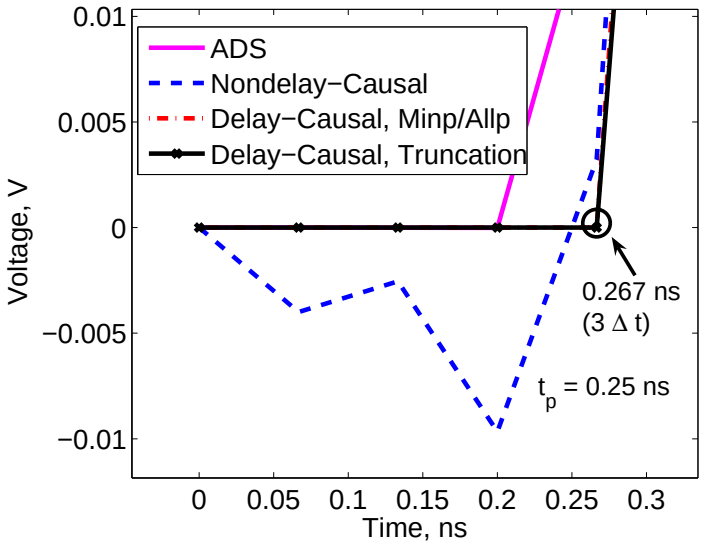



(b) FTs of the impulse responses in Figure 75(a).

**Figure 75. Comparison of transfer impulse responses and their transforms using truncation-based ('Truncation') and minimum-phase-based ('Minp/Allp') delay-causal techniques for a causal data.**




(a) Comparison of step responses from truncation-based technique and minimum-phase-based technique.




(b) Comparison of step responses from truncation-based technique and minimum-phase-based technique for $t \leq t_p$.

**Figure 76. Comparison of step responses obtained from different approaches with that from ADS.**

$\tilde{h}(t)$ (see Figure 75(b)). This disparity manifests in the incorrect final values of the step responses at $p_2$ (see Figure 76(a)) obtained from the corresponding impulse responses. From Figure 76(a-b), it can be observed that the minimum-phase-based delay-causal results are accurate. Also, it is to be noted that the ADS transient results do not capture the propagation delay in the line: the step response at the far end of the line from ADS is nonzero for $t < t_p$. However, in the the minimum-phase-based transient results, the propagation delay is correctly captured. The accuracy of the truncation-based delay-causal simulation can, however, be improved in this case, as $\hat{h}(t)$ converges to a delay-causal $\tilde{h}(t)$ with increasing $f_c$ (see Figure 77(a), case w/ $f_c = 8$ GHz). It has been observed for causal frequency-domain data nondelay-causal impulse response, $\hat{h}(t)$, converges to a delay-causal impulse response, $\tilde{h}(t)$, with increasing $f_c$. Therefore, accuracy of the truncation-based delay-causal simulation can be improved by the increasing the bandwidth of the data provided the data are already causal. In Figures 78 and 79, the convergence of the nondelay-causal step response to a delay-causal step response with increasing $f_c$ is shown. Also, from these two figures, it can also be observed that the step response from minimum-phase-based $\tilde{h}(t)$ is accurate (in this case, exact), as it is the converged nondelay-causal step response, obtained with $f_c = 8$ GHz. However, it can observed from Figure 79 that the propagation delay of 0.25 ns is not accurately captured in the minimum-phase-based technique: from Figure 79, this delay is approximately 0.215 ns. The reason for this disparity has to do with the bandwidth ($f_c = 7$ GHz) of the data available for the simulation. The delay 0.215 ns is approximately three times the time step ($\Delta t = 1/(2 \times 7 \text{ GHz})$) of the transient simulation.

2. $\breve{H}(f) \neq H(f)$, no $W(f)$ (Noncausal data, w/o windowing, see Figures 80-81): The comparison was performed for a lossy transmission line ($f_c = 10$ GHz, $t_p = 3$ ns) (see Figure 80) characterized by noncausal data: S-parameters are

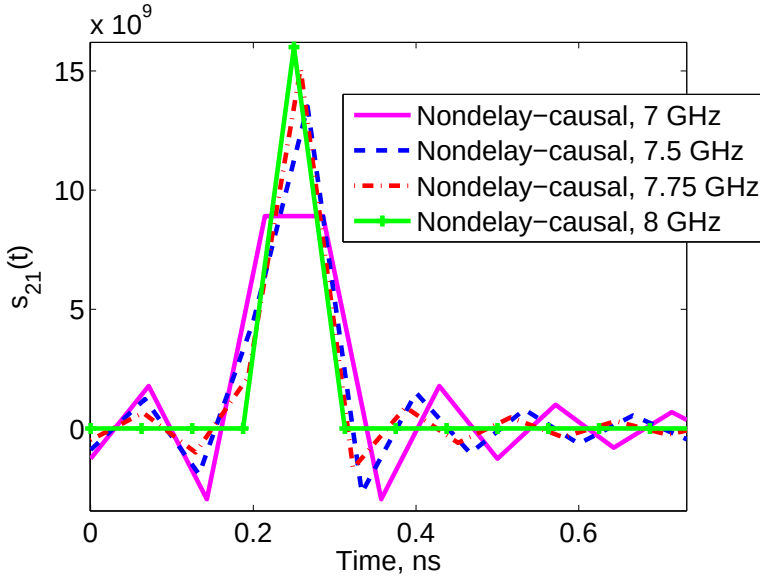



**Figure 77. Convergence of nondelay-causal impulse response, $\hat{h}(t)$, to a delay-causal impulse response, $h(t)$, with increase in bandwidth.**

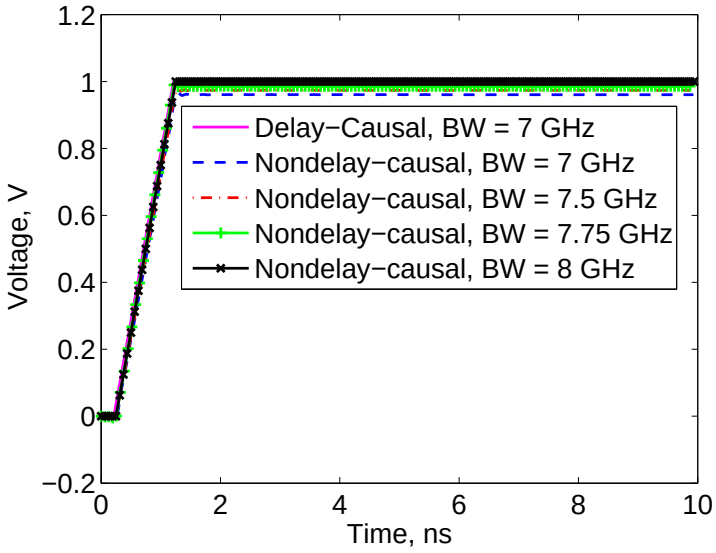
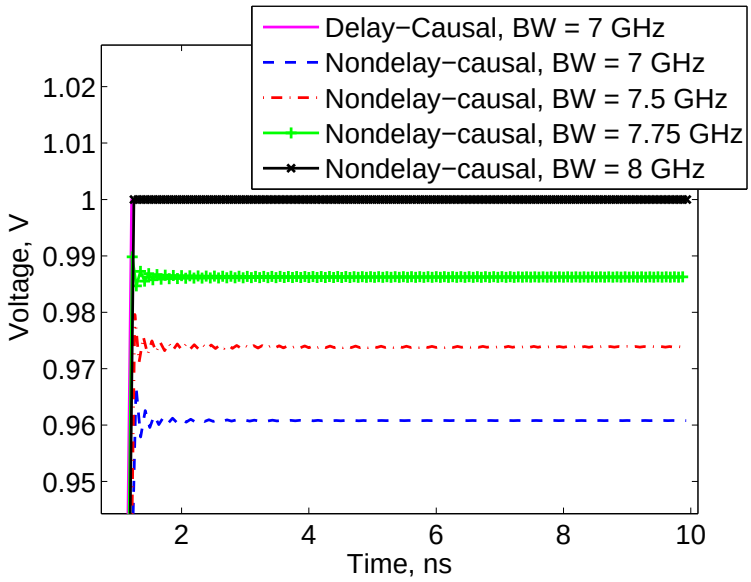
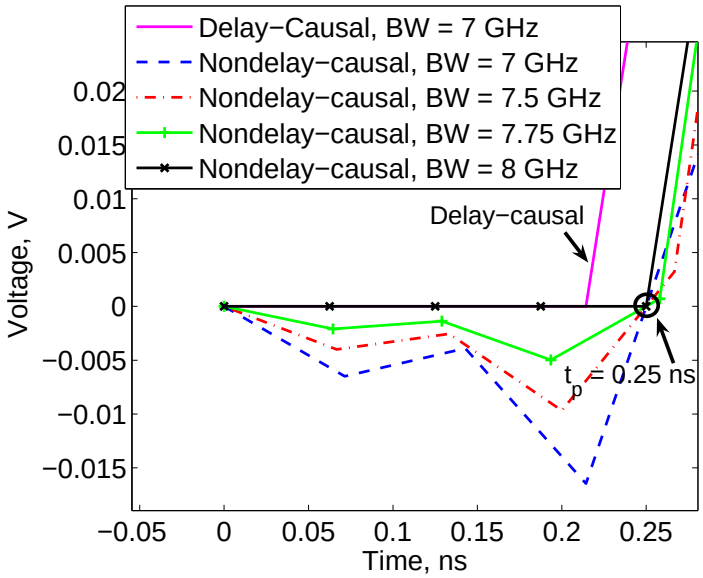



**Figure 78. Comparison of step responses obtained with nondelay-causal impulse responses of increasing bandwidth.**




(a) Convergence of DC level of the step response with increasing bandwidth




(b) Convergence of nondelay causal step response to a delay-causal step response with increasing bandwidth.

**Figure 79. Convergence of nondelay-causal step response to a delay-causal step response with increase in bandwidth.**

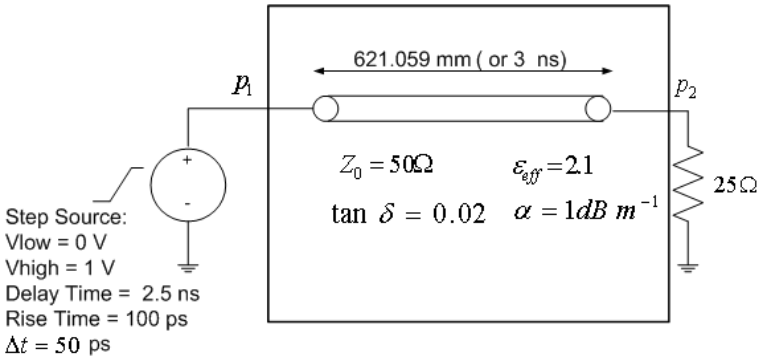



**Figure 80. Test setup of a step response of a lossy transmission line, $t_p = 3$ ns.**

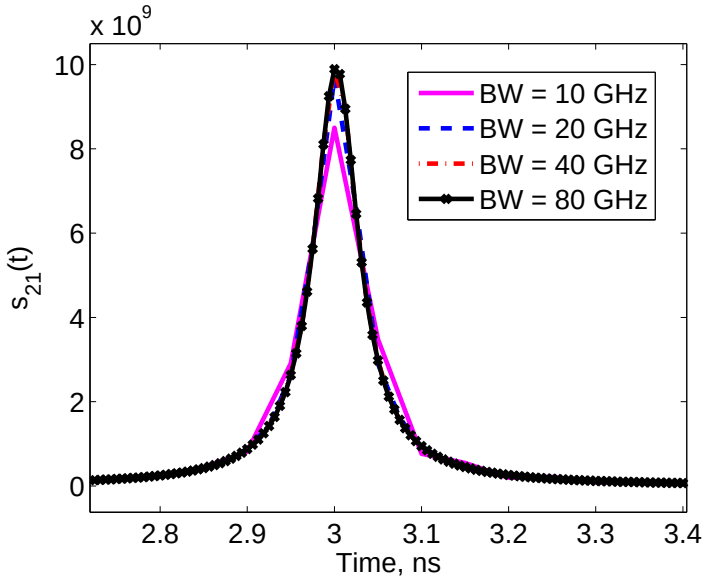



**Figure 81. Nonconvergence of nondelay-causal $\hat{h}(t)$ to a delay-causal response with increase in $f_c$.**

obtained from a noncausal circuit model that assumes constant $L(f)$ and $C(f)$ but frequency-dependent $R(f)$ ($\propto \sqrt{(f)}$) and $G(f)$ ($\propto f$). Because of the noncausal data, $\tilde{h}(t)$ is nondelay-causal and does not become delay-causal with increase in $f_c$ (see Figure 81). Since an appreciable part of $\hat{h}(t)$ is present for $t < t_p$, truncation results in a significant disparity in the DC levels of the step response (see Figure 82). This disparity cannot be alleviated with increasing $f_c$, as $\hat{h}(t)$ does not become delay-causal with increase in $f_c$ (see Figure 81). On the other hand, the minimum-phase-based technique results in a reasonably accurate result (see Figure 82). It is to be noted that, unlike

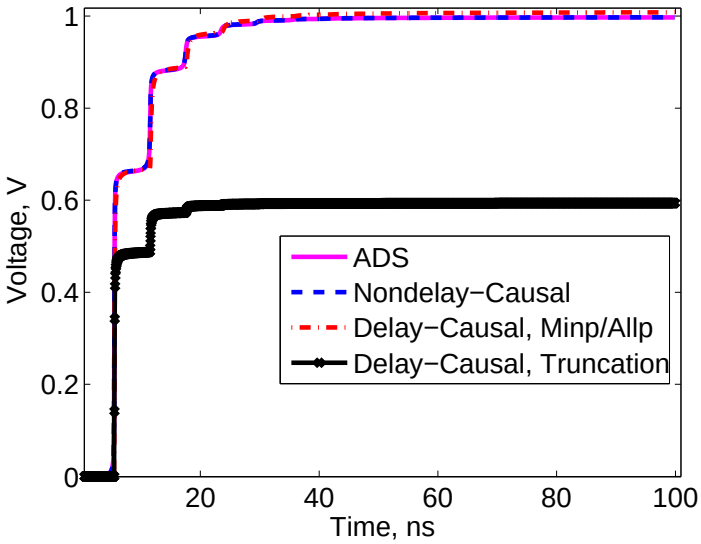
the transient results from ADS, the transient results from the minimum-phase-based technique captures the propagation delay accurately (see Figure 82(b)).

3. $\breve{H}(f) \neq H(f)$, $W(f)$ is present (Noncausal data, w/ windowing, See Figure 83-85):

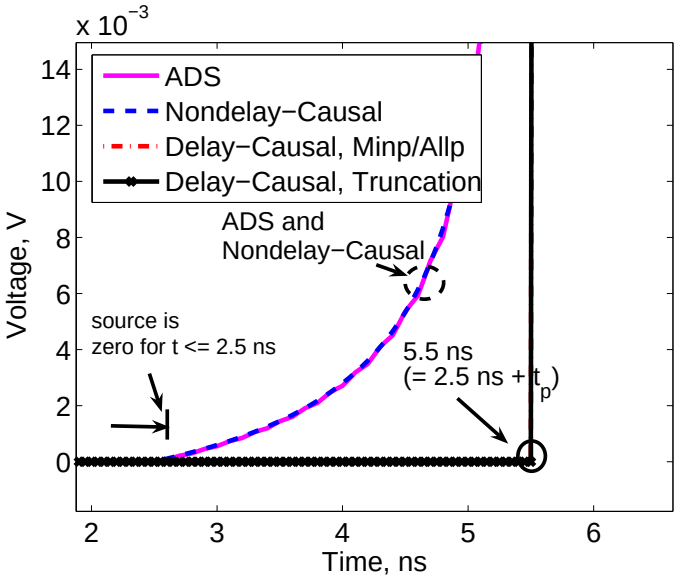
   When the windowing is applied, $\hat{h}(t)$ becomes broader (with center at $t_p$) and hence more nondelay-causal (see Figure 83). Note with windowing, low-frequency values of the transient response should not be affected much [88], particularly not the DC value (as W(0) = 1). However, when $\hat{h}(t)$ becomes more nondelay-causal (because of windowing), the truncation results in a bigger loss in the DC levels of the step response (see Figure 84, DC level drops by $\approx$ 120 mV). On the other hand, the results from minimum-phase-based technique are reasonably accurate (see Figure 85: DC values are preserved, and step response becomes more smooth when windowing becomes stronger).

## 9.6 Summary

Numerical convolution-based approach is common for transient simulation of interconnects characterized by band-limited data. When the data are band limited, the propagation delay is not accurately captured in the simulation. Common techniques that enforce propagation delay either truncate the nondelay-causal part of impulse response or construct a minimum-phase-reconstructed impulse response and shift this response by the delay. Both the techniques affect the accuracy of the simulation. In this paper, the effects of truncation-based and minimum-phase-based techniques on the accuracy of transient simulation are compared in the presence of delay-causality violations. It has been found that the truncation-based technique does not preserve the energy of the port-to-port frequency response and can affect the accuracy significantly, while the minimum-phase technique preserves the energy and does not have

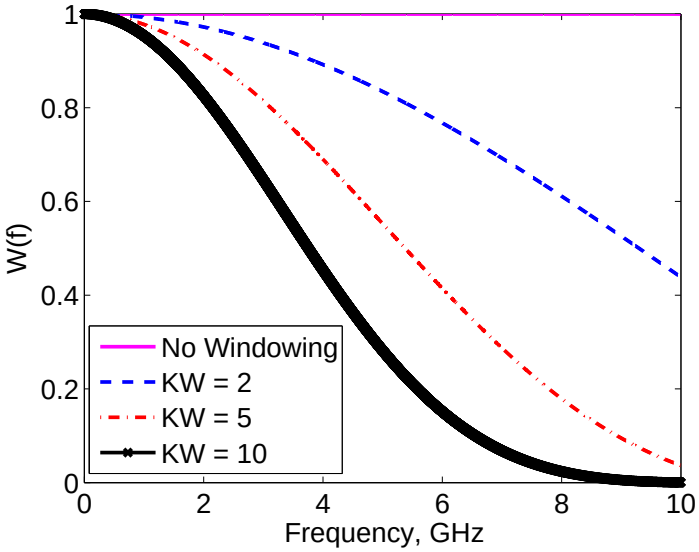



(a) Comparison of the DC level of the step responses.

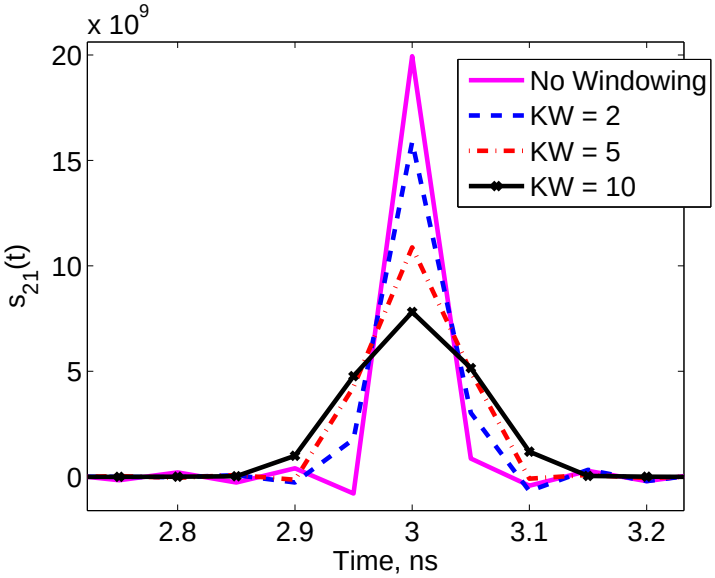



(b) Comparison of the delay-causal or nondelay-causal nature of the step responses.

**Figure 82. Comparison of step responses from truncation-based technique and minimum-phase-based technique for the test setup in Figure 80.**

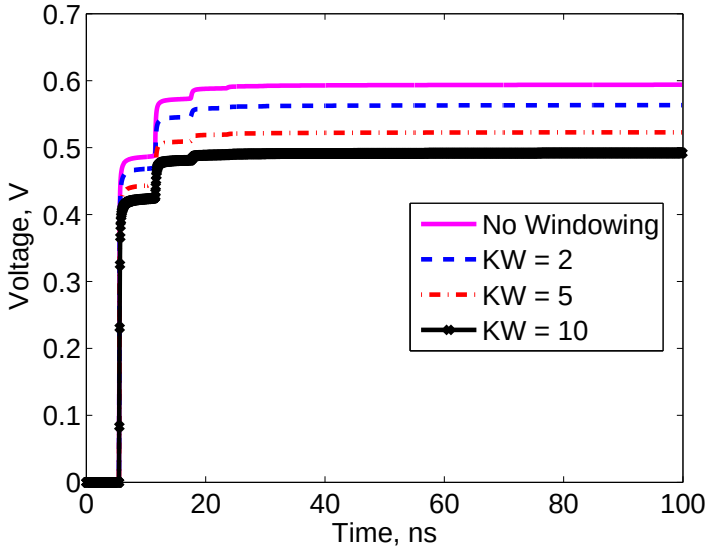



(a) Frequency-domain Kaiser windowing of varying shape parameters.




(b) Stronger (Kaiser) windowing (indicated by larger $KW$) makes the impulse response more broad at $t = 3$ ns.

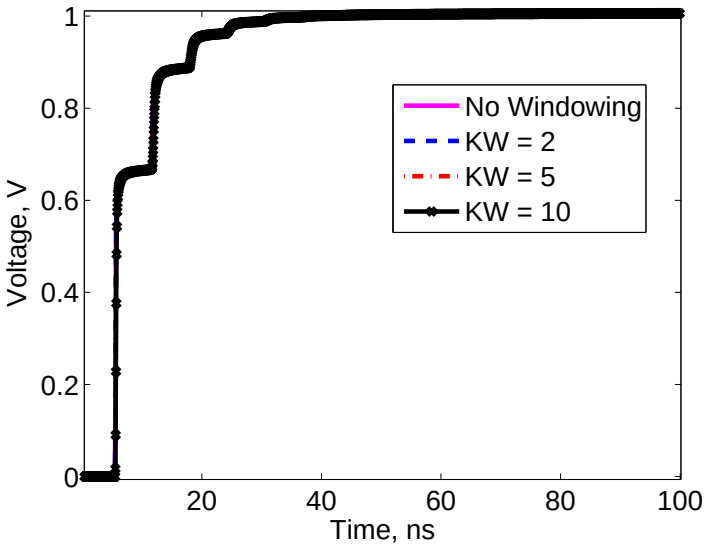
**Figure 83. Frequency-domain windowing makes the impulse response more nondelay-causal.**



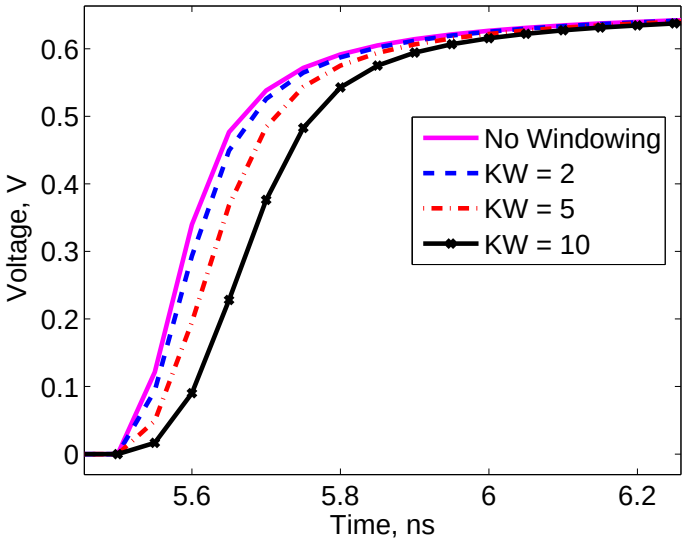

(a) Artificial loss of DC level using the truncation-based technique in the presence of windowing.

**Figure 84. Incorrect performance of the truncation-based technique in the presence of frequency-domain windowing.**

this inaccuracy problem. Frequency-domain windowing can worsen delay-causality violations. Minimum-phase-based technique has been applied with reasonable success for correcting delay-causality violations because of band-limited data, noncausal data, and windowing.




(a) Comparison of step responses obtained minimum-phase-based technique in the presence of windowing.




(b) The effect of slowing step response because of stronger windowing is captured using a minimum-phase-based technique.

**Figure 85. Reasonable accurate transient simulation using the minimum-phase-based technique in the presence of windowing.**

# CHAPTER 10

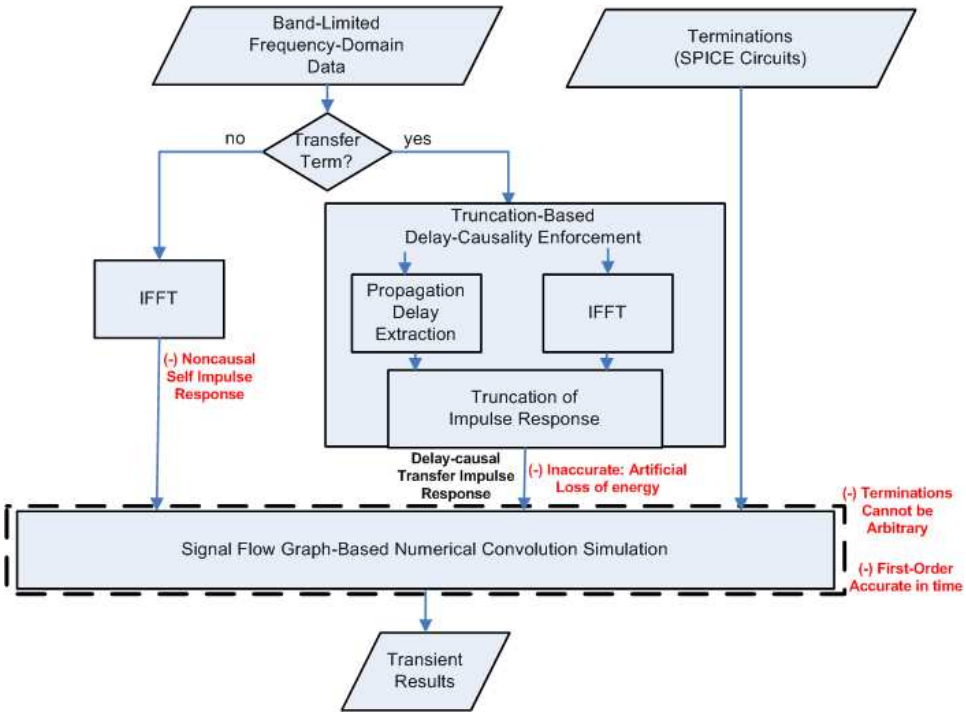
# GENERALIZED LINEAR PHASE CONDITION AND HANDLING ARBITRARY TERMINATIONS THROUGH A MODIFIED NODAL ANALYSIS FRAMEWORK

## 10.1 Introduction

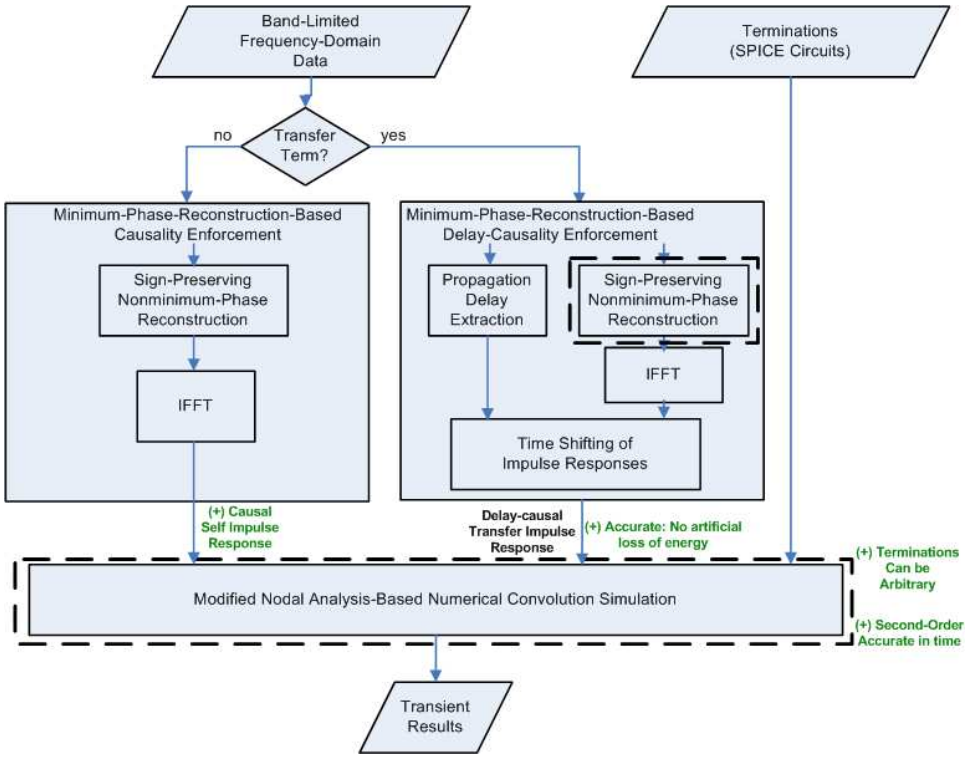
In Chapter 9, it was established that delay-causality enforcement of transfer frequency responses through a minimum-phase/all-pass decomposition can be more accurate than through a truncation-based approach. In this chapter, this decomposition is extended to reconstruct responses that cannot be faithfully reconstructed. In this chapter, it is first shown that the functional form of the all-pass component proposed in Chapter 9 may not capture the leading negative sign in frequency responses when they are reconstructed. As a result, in this chapter, a new functional form for the all-pass component is proposed. This form is needed for capturing the leading negative sign in frequency responses during their reconstruction. Also, in this chapter, it has been shown that with a signal flow graph-based approach [55], [56], [57] for convolution-based transient simulation, it is difficult to handle arbitrary port terminations. Subsequently, in this chapter, a new transient simulation algorithm based on a modified nodal analysis framework has been proposed. The advantage of the this new framework is that arbitrary port terminations can be handled with ease. Numerical results demonstrating the accuracy and the capability of the proposed procedure have been presented. The focus of this chapter is also described in Figure 86.

## 10.2 Short Background

In many applications, only the band-limited (b.l.) frequency-domain (f.d.) data (e.g., S-, Y-, Z-parameters) of an interconnect (e.g. a lossy transmission line) are known. The objective is to perform an accurate transient simulation of the multiport b.l.f.d. data with the port terminations. Such a simulation is useful for studying pulse




(a) Prior approach [55], [56], [57].




(b) Proposed approach in this dissertation.

**Figure 86. Comparison of the prior and proposed approach in numerical-convolution-based causal transient simulation of band-limited data. The focus of this chapter is the region marked within the dashed rectangle.**

propagation in a transmission line or for computing crosstalk in coupled transmission lines when only the S-parameters of the lines are known up to a given frequency. In such a simulation, it is of interest 1) to capture the propagation delay through the interconnects and 2) to conveniently handle arbitrary port terminations.

Most of the prior work in the transient simulation of b.l.f.d. data employ a recursive-convolution-based approach [58], [60], [91], [53], [92], [61], [59]. However, this approach can become computationally exorbitant for large number of ports, $N_p$, and/or for large number of poles, $N_{pl}$ [60]. This computational inefficiency is mainly because of the rational-function fitting procedure required in this approach. Remaining prior work is based on a numerical-convolution-based approach [62], [63], [64], [65], [55], [56], [57]. This approach does not suffer from the computational inefficiency associated with the rational-function fitting step. Most of the prior work using the numerical-convolution-based approach do not capture the port-to-port propagation delays in the transient simulation when only the b.l.f.d. data are known about interconnects [62]- [65]. In the prior work that does capture the propagation delays when only the b.l.f.d. data are known, namely [55], [56], arbitrary equivalent circuits for port terminations cannot be conveniently handled. In this chapter, a numerical-convolution-based approach is proposed that not only captures port-to-port propagation delays but also conveniently handles arbitrary port terminations. Such a handling is accomplished by integrating the numerical convolution in a modified-nodal analysis framework, unlike [55], [56]. The proposed formulation uses a minimum-phase-based reconstruction approach with a sign-preservation term. This extra term, which is missing in [55], [56], is essential to obtaining accurate transient results in certain examples.

*The contribution of this chapter*[1] *are the following:*

[1] **S. N. Lalgudi**, E. Engin, G. Casinovi, and M. Swaminathan , "Accurate Transient Simulation of Interconnects Characterized by Band-Limited Data With Propagation Delay Enforcement in a Modified Nodal Analysis Framework," Accepted for Publication in *IEEE Trans. on Electromagnetic Compatibility*, July 2007.

1. *Numerical-convolution-based delay-causal transient simulation of interconnects characterized by multiport band-limited data that can also conveniently handle arbitrary port terminations.*

2. *Sign-preserving minimum-phase/all-pass decomposition for the delay-causality enforcement.*

The rest of this chapter is organized as follows: In Section 10.3, the delay-causality problem with b.l.f.d data has been mathematically formulated. Also, in this section, the procedure to obtain a delay-causal impulse response from the b.l. data using the proposed form for the all-pass component has been described. In Section 10.4, the numerical convolution-based delay-causal transient simulation procedure has been explained. In Section 10.5, the proposed procedure to handle terminations in an MNA framework has been described. In Section 10.6, simulation results demonstrating the accuracy of the proposed decomposition and of the proposed transient simulation procedure have been presented. Finally, in Section 10.7, the conclusions of this chapter have been presented.

## 10.3 Delay-Causality Problem

The delay-causality problem solved in this work, as well as in [55], [56], can be mathematically stated as follows: Consider a linear time-invariant passive system (the black box in Figure 87) with an impulse response $h(t)$ and a propagation delay $t_p$. The impulse response $h(t)$ is delay-causal. Let this system be fed by a time-domain signal $x(t)$, and let the time-domain response at the output be $y(t)$. The objective is to find an approximate delay-causal output, $\tilde{y}(t)$, given $x(t)$ and the frequency response (in terms of Y-, Z-, S-parameters), $H(\omega)$, of the system at uniformly-spaced frequency intervals between 0 to $f_c$, where $f_c$ is some high-enough frequency.

Since $x(t)$ and $y(t)$ are related through convolution, $\tilde{y}(t)$ can be computed if an approximate delay-causal impulse response, $\tilde{h}(t)$, can be found (see [55], [56]). If $\hat{h}\,(t)$

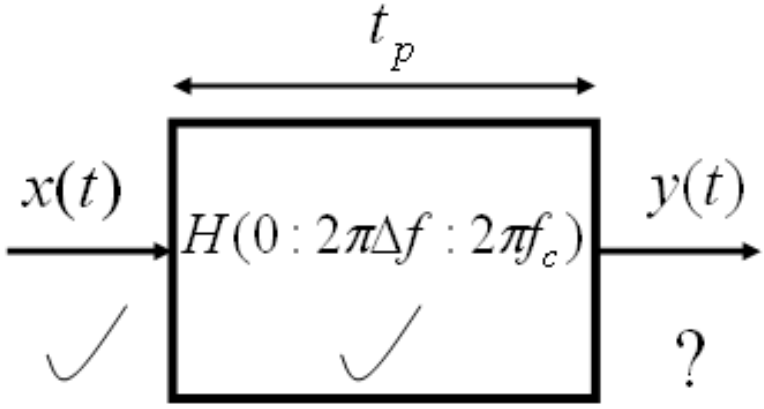



**Figure 87. Definition of the causality problem: Given $x(t)$ and the band-limited and sampled frequency data, $H(\omega)$, of a passive system with a propagation delay, $t_p$, find the output $y(t)$ such that $t_p$ is strictly enforced in $y(t)$; $\Delta f$ is frequency step of the sampled data, and $f_c$ is some high-enough frequency up to which the data are known. A tick mark indicates a known (or given) quantity, and the question mark indicates an unknown quantity to be computed.**

denotes the inverse fourier transform (IFT) of the b.l. response $H(0:2\pi\Delta f:2\pi f_c)$, where $\Delta f$ is the frequency step, then $\hat{h}\,(t)$ is not the preferred solution, as $\hat{h}\,(t)$ is not delay-causal [55], [56]. This is because when $f_c$ is finite (equivalent to multiplying the infinite frequency response $H(0:2\pi\Delta f:\infty)$ by a gate function of width $f_c$), $\hat{h}\,(t)$ is actually the convolution of a time-domain sinc function, which is noncausal, and $h(t)$, which is delay-causal, Therefore, $\hat{h}\,(t)$ can be nondelay-causal and may be noncausal too. Then, the objective is to find a $\tilde{h}(t)$ that approximates $h(t)$ from only $H(0:2\pi\Delta f:2\pi f_c)$.

In the rest of this chapter, the procedure in [55], [56] to obtain the delay-causal impulse response from b.l. data has been briefly explained, followed by the description of a possible limitation of this procedure in preserving the sign of the original frequency response. Next, a new decomposition for the frequency response that removes this limitation has been proposed.

### 10.3.1 Delay-Causal Impulse Response using Linear-Phase Condition

In [55], [56], a decomposition procedure for $H(\omega)$ that results in a $\tilde{h}(t)$ has been presented. This decomposition is given as

$$H(\omega) = H_{\min}(\omega)\, e^{-j\omega t_p}, \tag{86}$$

where $H_{\min}(\omega)$ is the minimum-phase component (see [89], [87]). The function $H_{\min}(\omega)$ is the delayless part of $H(\omega)$ and models effects because of attenuation and dispersion. The term $e^{-j\omega t_p}$ is the all-pass component, which has been used to model the phase using a linear-phase condition. The term $t_p$ here refers to the propagation delay if $H(\omega)$ was lossless. For a lossy transmission line of length $l$, $t_p$ would mean the propagation delay of a lossless line of the same length and is calculated by the value of the propagation delay at $\omega = \infty$ [90].

The IFT of the b.l. response $H_{\min}(0 : \omega_c)$, $\hat{h}_{\min}(t)$, is causal (see [89]). Therefore, the IFT of the R.H.S. of (86), $\hat{h}_{\min}(t - t_p)$, is delay-causal. Therefore, $\tilde{h}(t)$ is chosen as

$$\tilde{h}(t) = \hat{h}_{\min}(t - t_p). \tag{87}$$

The component $H_{\min}(\omega)$ in (86) is computed from the magnitude of $H(\omega)$:

$$|H_{\min}(\omega)| = |H(\omega)|; \tag{88}$$

$$\arg[H_{\min}(\omega)] = -\mathrm{HT}\{\ln|H(\omega)|\}. \tag{89}$$

In (88) and (89), $|x|$ stands for the magnitude of $x$, $\arg[x]$ is the principal argument of complex number $x$, and $\mathrm{HT}\{\mathrm{x}\}$ is the Hilbert transform [89] of $x$. Using a discrete Hilbert transform [89], (89) can be rewritten as

$$\arg[H_{\min}(\omega)] = -\frac{1}{2\pi}\mathcal{P}\int_{\theta=-\pi}^{\pi} \ln|H(\theta)| \cot\left(\frac{\omega-\theta}{2}\right) d\theta, \tag{90}$$

where $\mathcal{P}$ denotes the Cauchy principal value of the integral that follows. The propagation delay, $t_p$, is extracted using the procedure described in Chapter 9 (Section 9.4.1).

### 10.3.2 A Limitation of Linear-Phase Condition

The procedure thus far described works as long as $H(\omega)$ can be decomposed according to the functional form described in (86). However, when $H(\omega)$ has a constant negative sign, a simple example is $H(\omega) = -1$, the decomposition in (86) is not sufficient. The off-diagonal terms of the admittance matrix of a resistive circuit has a form $H(\omega) = -g$, where $g > 0$ is the conductance between two different ports.

To see the insufficiency of (86), $H(\omega) = -g$ is reconstructed using (86). From (88), $|H_{\min}(\omega)| = g$. From the property of Hilbert transforms, the Hilbert transform of a constant is zero [93]. This can also be proven from (90) by deducing that for a constant $H(\omega)$, the integrand is an odd function. Therefore, the integration result is zero. Using this fact in (89), $\arg[H_{\min}(\omega)] = 0$. Therefore, $H_{\min}(\omega) = g$.

Since in a resistive circuit, there is no propagation delay between ports, $t_p = 0$. Therefore, the exponential term in (86), call it $H_{\text{ap}}(\omega)$, is 1.

From $H_{\min}(\omega) = g$ and $H_{\text{ap}}(\omega) = 1$, the original response $H(\omega) = -g$ is reconstructed as only $g$ using the decomposition in (86)! In fact, since only the magnitude of $H(\omega)$ is used to compute $H_{\min}(\omega)$, all frequency responses of the form $H(\omega) = ge^{j\theta}$ will be reconstructed as just $g$, where $\theta$ is a constant real number. This disparity in the phase between the original frequency response and the reconstructed frequency response could affect the accuracy of the transient results, as will be shown in Section 10.6.

### 10.3.3 Delay-Causal Impulse Response using Generalized-Linear Phase Condition

To account for a constant phase term in the frequency response, the form of the decomposition in (86) is modified as

$$H(\omega) = H_{\min}(\omega)\, e^{-j\omega t_p + j\theta}. \tag{91}$$

For the example $H(\omega) = -g$, $\theta = \pm\pi$. Therefore, using (91), $H(\omega) = -g$ can be reconstructed from $H_{\min}(\omega) = g$, $t_p = 0$, and $\theta = \pi$. Therefore, the proposed all-pass

component form is $e^{-j\omega t_p + j\theta}$. The resulting condition on the phase of the all-pass component is referred to as the generalized linear-phase condition, a condition used to denote a generalized linear-phase system (see [89], pp. 295).

The constant phase $\theta$ in (91) can be computed numerically from the frequency data by 1) equating the phases of the L.H.S. and the R.H.S. of (91) and 2) solving for $\theta$ from the resulting equation, which can be written as

$$\theta = \arg\left[H\left(\omega\right)\right] - \arg\left[H_{\min}\left(\omega\right)\right] - \arg\left[e^{-j\omega t_p}\right]. \tag{92}$$

The phase $\theta$ from (92) can be computed by obtaining the R.H.S. at any $\omega$ or by calculating the average of the R.H.S.'s for all $\omega$'s. However, it has to be noted that $t_p$ is only computed in the average sense (see Section 10.3.1) and hence can contribute to some inaccuracy while calculating the term $\omega t_p$ in (92). This inaccuracy issue can be avoided if $\theta$ is computed by obtaining the R.H.S at $\omega = 0$. However, computing $\theta$ at $\omega = 0$ is not reliable for the following reason: At $\omega = 0$, the magnitude of the transfer response can be zero, making the angle of the response zero too at $\omega = 0$. Such a case arises in coupled transmission lines. Also, the angle of the minimum-phase response at $\omega = 0$ is always zero. For $\omega = 0$, the integrand in (90) is an odd function of $\theta$. Therefore, at $\omega = 0$, $\theta$ can be computed to be zero. Therefore, the phase $\theta$ is computed near $\omega = 0$. If the angle of the original frequency response or of its minimum-phase component is discontinuous near $\omega = 0$, then this angle is computed in the asymptotic sense (value of the angle as $\omega \to 0$).

Irrespective of the $\omega$ at which $\theta$ is computed, there are some restrictions on the values $\theta$ can take: The term $e^{j\theta}$ in (91) introduces a constant phase change to the rest of the response for all frequencies including $\omega = 0$. Since $H(0)$ and $H_{\min}(0)$ are both real, the term $e^{j\theta}$ can only be a real number. Therefore, the phase angle $\theta$ can take values among 0, $\pi$, and $-\pi$ radians. These values mean that the term $e^{j\theta}$ at the most can result in and account for a sign change.

With the proposed decomposition in (91), the impulse response $\tilde{h}(t)$ in (87) is

computed differently, as the IFT of the R.H.S. of (91), i.e.,

$$\tilde{h}(t) = \hat{h}_{\min,\theta}(t - t_p), \tag{93}$$

where $\hat{h}_{\min,\theta}(t)$ is the b.l. IFT of the product $H_{\min}(\omega) e^{j\theta}$.

## 10.4 Numerical Convolution-based Delay-Causal Transient Simulation

Using (93), all impulse responses between two different ports are obtained as delay-causal impulse responses. However, impulse responses between same ports (i.e., $s_{ii}(t)$, $y_{ii}(t)$, etc) are obtained as the IFT of the corresponding frequency responses, as is being done in [55], [56]. This different treatment to self terms is because of the following considerations: 1) Self impulse responses represent reflection (or return loss) characteristics at a port because of an excitation at the same port. As there is no delay between the same ports, propagation delays for self terms are made zero. In case of multiple delays (happens when characteristic impedance is not equal to the reference impedance), the smallest of them is zero. 2) Port-to-port frequency responses between same ports are considered as minimum phase [94], and minimum-phase frequency responses have a causal time-domain response [89]. Therefore, self impulse responses are automatically delay-causal with a delay of zero.

Once multiport impulse responses are known, the transient simulation involves computing port voltages given the equivalent circuits of port terminations. For a numerically robust transient simulation, the f.d. data are expressed as S-parameters [63]. The transient simulation requires solving the convolution equations relating the port quantities, such as the incident and the reflected waves, with the equations describing the termination conditions. In the rest of this section, the convolution equations are derived.

Let $\overline{\overline{S}}(\omega) \in \mathbf{C}^{N_p \times N_p}$ be the multiport S-parameter. Then, $\overline{\overline{S}}(\omega)$ can be written as

$$\overline{\overline{S}}(\omega) = \overline{\overline{S}}(\infty) + \overline{\overline{\hat{S}}}(\omega), \tag{94}$$

where $\overline{\overline{S}}(\infty)$ is $\overline{\overline{S}}(\omega)$ at $\omega = \infty$ and is because of the direct coupling between input and output ports, and $\overline{\overline{\hat{S}}}(\omega)$ is the remaining part of $\overline{\overline{S}}(\omega)$. If $\bar{A}(\omega) \in \mathbf{C}^{N_p \times 1}$ and $\bar{B}(\omega) \in \mathbf{C}^{N_p \times 1}$, respectively, are the vector of incident and reflected waves [95], then $\bar{B}(\omega) = \overline{\overline{S}}(\omega)\bar{A}(\omega)$, which in time domain becomes

$$\overline{b}(t) = \overline{\overline{s}}(t) * \overline{a}(t). \tag{95}$$

In (95), $\overline{\overline{s}}(t)$, $\overline{a}(t)$, and $\overline{b}(t)$ are the IFT's of $\overline{\overline{S}}(\omega)$, $\bar{A}(\omega)$, and $\bar{B}(\omega)$, respectively. The symbol '*' in (95) denotes a linear convolution [89] and is defined as

$$y(t) = h(t) * x(t) = \int_{\tau=0}^{t} h(t-\tau)x(\tau)\,d\tau. \tag{96}$$

If $h(t)$ in (96) does not have any impulses, then the continuous integration in (96) can be discretized using a (right) rectangular integration rule as

$$y(t) \approx \sum_{m=1}^{n} h((n-m)\Delta t)\,x(m\Delta t)\,\Delta t + O(\Delta t), \tag{97}$$

where $\Delta t$ is the time step, and $O(\Delta t)$ denotes the first-order accuracy of the integration rule. Defining $\overline{\overline{\hat{s}}}(t)$ to be the IFT of $\overline{\overline{\hat{S}}}(\omega)$ and $\delta(t)$ to be the Dirac-Delta function, $\overline{\overline{s}}(t)$ can be written as

$$\overline{\overline{s}}(t) = \overline{\overline{S}}(\infty)\,\delta(t) + \overline{\overline{\hat{s}}}(t). \tag{98}$$

Making use of (98), (96), and (97), Equation (95) can be written as

$$\overline{b}(t) \approx \overline{\overline{S}}(\infty)\,\overline{a}(t) + \sum_{m=1}^{n} \overline{\overline{\hat{s}}}((n-m)\Delta t)\,\overline{a}(m\Delta t)\Delta t. \tag{99}$$

When the $n$th term in the summation in (99) is separated and combined with the first term of the R.H.S. of (99), then the resulting equation can be rewritten as

$$-\left[\overline{\overline{S}}(\infty) + \overline{\overline{\hat{s}}}(0)\,\Delta t\right]\bar{a}(t) + \bar{b}(t) = \bar{h}(t), \tag{100}$$

where

$$\bar{h}(t) = \sum_{m=1}^{n-1} \overline{\overline{\hat{s}}}((n-m)\Delta t)\,\bar{a}(m\Delta t)\,\Delta t. \tag{101}$$

From (101), it can be observed that $\bar{h}(t)$ depends only on the known values of $\bar{a}$; hence, the R.H.S. of (100) is known. However, $\bar{a}\,(t)$ and $\bar{b}\,(t)$ in (100) are still not known. Therefore, (100) constitutes a set of $N_p$ equations with $2N_p$ unknowns (both $\bar{a}(t)$ and $\bar{b}(t)$). The system in (100) has be solved together with the equations describing the terminations.

## 10.5 Handling Terminations

In this section, the procedure to handle port terminations in SFG-based approaches [55], [56], [96] has been briefly explained, followed by a description of its limitation to handle complicated terminations. Next, the MNA-based convolution simulation that handles terminations without the limitations in an SFG-based approach has been proposed.

### 10.5.1 Handling Terminations in an SFG-based Approach

Since both $\bar{a}\,(n\Delta t)$ and $\bar{b}\,(n\Delta t)$ are still not known in (100), at least another $N_p$ equations are needed to compute them. The additional $N_p$ equations are obtained by relating $\overline{a}\,(t)$ and $\overline{b}\,(t)$ through the termination conditions [96]:

$$\overline{a}\,(t) = \overline{\overline{\Gamma}}\,(t)\,\overline{b}\,(t) + \overline{\overline{\mathrm{T}}}\,(t)\,\overline{g}\,(t)\,. \tag{102}$$

In (102), $\overline{\overline{\Gamma}}\,(t) \in \mathbf{R}^{N_p \times N_p}$ and $\overline{\overline{\mathrm{T}}}\,(t) \in \mathbf{R}^{N_p \times N_p}$ are the diagonal matrices of the reflection and the transmission coefficients at ports at time $t$, respectively. The vector $\overline{g}\,(t) \in \mathbf{R}^{N_p \times 1}$ is a function of the excitations at ports and is known at time $t$. The port quantities $\overline{a}\,(t)$ and $\overline{b}\,(t)$ can now be obtained by solving (100) together with (102). Let $N_n$ denote the total number of nodes in the network, and let the first $N_p$ nodes correspond to the $N_p$ ports. If $\overline{v}\,(t) \in \mathbf{R}^{N_n \times 1}$ denotes the vector of node voltages, then the port voltages can be computed as

$$\overline{v}_{1:N_p}\,(t) = \overline{a}\,(t) + \overline{b}\,(t)\,. \tag{103}$$

In [55], [56], the port voltages at every time step are computed by solving (100), (102), and (103) . The disadvantage of such a computation is that the matrices $\overline{\overline{\Gamma}}(t)$ and $\overline{\overline{\mathrm{T}}}(t)$ in (102) are difficult to compute when terminations have complicated equivalent circuits, as computing these matrices require computing driving point impedances looking away from the ports.

### 10.5.2 Handling Terminations in an MNA-based Approach

This difficulty can be avoided if the termination conditions in (102) are alternatively enforced through a modified nodal analysis formulation. If $\overline{i}(t) \in \mathbf{R}^{N_p \times 1}$ is the vector of currents entering the ports, then the MNA of the whole network (multiport network + rest of the network) yields the following system of equations:

$$\overline{\overline{C}}\,\overset{\bullet}{\overline{x}}(t) + \overline{\overline{G}}\overline{x}(t) + \begin{bmatrix} \overline{i}(t) \\ \overline{0}_{N_{\mathrm{mna}}-N_p} \end{bmatrix} = \overline{r}(t), \tag{104}$$

where $\overset{\bullet}{\overline{x}}(t) = \frac{d\overline{x}(t)}{dt}$, $\overline{x}(t) \in \mathbf{R}^{N_{\mathrm{mna}} \times 1}$ is the vector of unknown variables in an MNA approach, and $\overline{r}(t) \in \mathbf{R}^{N_{\mathrm{mna}} \times 1}$ is a vector describing the current and voltage sources in the whole network. The quantities $\overline{\overline{C}} \in \mathbf{R}^{N_{\mathrm{mna}} \times N_{\mathrm{mna}}}$ and $\overline{\overline{G}} \in \mathbf{R}^{N_{\mathrm{mna}} \times N_{\mathrm{mna}}}$, $\overline{r}(t) \in \mathbf{R}^{N_{\mathrm{mna}} \times 1}$ have the same definitions as in the MNA approach, and $N_{\mathrm{mna}} = N_n + N_{v_s} + N_L$. The symbol $N_{v_s}$ denotes the total number of voltage sources in the network, and the symbol $N_L$ denotes the total number of inductors in the network. In (104), the symbol $\overline{0}_k$ denotes a column vector of zeros with $k$ rows.

Since $\overline{i}(t)$ in (104) is dependent on the f.d. data, the MNA system in (104) cannot be solved alone. Assuming all ports are referenced with respect to a characteristic admittance of $Y_0 \in \mathbf{R}$, the port currents can be expressed as

$$\overline{i}(t) = Y_0\left(\overline{a}(t) - \overline{b}(t)\right). \tag{105}$$

When $\overline{i}(t)$ in (105) is substituted in (104), the latter equation can be rewritten as

$$\overline{\overline{C}}\,\overset{\bullet}{\overline{x}}(t) + \overline{\overline{G}}\overline{x}(t) + Y_0 \begin{bmatrix} \overline{a}(t) \\ \overline{0}_{N_{\mathrm{mna}}-N_p} \end{bmatrix} - Y_0 \begin{bmatrix} \overline{b}(t) \\ \overline{0}_{N_{\mathrm{mna}}-N_p} \end{bmatrix} = \overline{r}(t). \tag{106}$$

To solve for all the node voltages including the port voltages, (106) is solved along with (100) and (103). The system combining these equations can be written as

$$\overline{\overline{W}}\,\overset{\bullet}{\overline{u}}\,(t) + \overline{\overline{V}}\,\overline{u}\,(t) = \overline{z}\,(t)\,, \tag{107}$$

where $\overline{u}\,(t) \in \mathbf{R}^{N_{\mathrm{mna}}+2N_p\times 1}$, $\overline{\overline{W}} \in \mathbf{R}^{N_{\mathrm{mna}}+2N_p\times N_{\mathrm{mna}}+2N_p}$, $\overline{\overline{V}} \in \mathbf{R}^{N_{\mathrm{mna}}+2N_p\times N_{\mathrm{mna}}+2N_p}$, and $\overline{z}\,(t) \in \mathbf{R}^{N_{\mathrm{mna}}+2N_p\times 1}$. These quantities are defined as follows:

$$\bar{u}\,(t) = \begin{bmatrix} \overline{x}\,(t) \\ \overline{a}\,(t) \\ \overline{b}\,(t) \end{bmatrix}. \tag{108}$$

$$\overline{\overline{W}} = \begin{bmatrix} \overline{\overline{C}} & \overline{\overline{0}}_{N_{mna}\times 2N_p} \\ \overline{\overline{0}}_{2N_p\times N_{mna}} & \overline{\overline{0}}_{2N_p\times 2N_p} \end{bmatrix}. \tag{109}$$

$$\overline{\overline{V}} = \begin{bmatrix} \overline{\overline{G}} & Y_0\overline{\overline{I}}_{N_p} & -Y_0\overline{\overline{I}}_{N_p} \\ \overline{\overline{0}}_{N_p\times N_{\mathrm{mna}}} & -\left(\overline{\overline{S}}\,(\infty) + \overline{\overline{\hat{s}}}\,(0)\,\Delta t\right) & \overline{\overline{I}}_{N_p} \\ \overline{\overline{I}}_{N_p} \quad \overline{\overline{0}}_{N_p\times N_{\mathrm{mna}}-N_p} & \overline{\overline{I}}_{N_p} & \overline{\overline{I}}_{N_p} \end{bmatrix}. \tag{110}$$

$$\bar{z}\,(t) = \begin{bmatrix} \overline{r}\,(t) \\ \overline{h}\,(t) \\ \overline{0}_{N_p} \end{bmatrix}. \tag{111}$$

In (109) and (110), the symbol $\overline{\overline{0}}_{m\times n}$ denotes a matrix of zeros with $m$ rows and $n$ columns, and $\overline{\overline{I}}_m$ denotes an identity matrix of size $m$. The unknown node voltages ($\overline{u}_{1:N_p}\,(t)$) can be computed from the solution of (107). The system (107) has the same form as the system most SPICE-like simulators (see [97], [44]) have. Therefore, numerical techniques to solve (107) are the same as those employed in SPICE-like simulators. With the formulation described thus far, any linear termination can be handled without having to compute the reflection or the transmission coefficients at ports.

The proposed formulation can also be extended to nonlinear terminations. It is to be noticed that when terminations are linear, (107) would represent a system of linear algebraic equations. This linear system of equations can be solved using linear matrix solution techniques. On the other hand, if terminations are nonlinear, Equation (104) (therefore, even (106)) would have nonlinear terms in addition to the existing terms, as part of the MNA of nonlinear elements. Equation (107) would therefore represent a system of nonlinear algebraic equations, which can be solved using the Newton-Raphson method [98].

The explicit splitting of $\overline{\overline{S}}(\omega)$ described in (94) can be avoided by dividing S-parameters by $\Delta t$ before computing impulse responses from them and using IFFT [89] to obtain the impulse responses: Defining $\overline{\overline{p}}(t)$ to be the IFFT [89] ($\neq$ IFT. Note IFFT and IFT results can differ by a factor of $\Delta t$) of $\frac{\overline{\overline{S}}(\omega)}{\Delta t}$, $\overline{\overline{p}}(t)$ can be expressed as

$$\overline{\overline{p}}(t) = \frac{\overline{\overline{S}}(\infty)}{\Delta t}\delta(t) + \overline{\overline{\hat{s}}}(t). \tag{112}$$

From (112), the following can be inferred:

$$\overline{\overline{S}}(\infty) + \overline{\overline{\hat{s}}}(0)\,\Delta t = \overline{\overline{p}}(0)\,\Delta t \tag{113}$$

and

$$\overline{\overline{\hat{s}}}(t \neq 0) = \overline{\overline{p}}(t \neq 0). \tag{114}$$

Using (113) and (114) in (100), it can be observed that (100) can be rewritten only in terms of $\overline{\overline{p}}(t)$, which is obtained without any splitting to $\overline{\overline{S}}(\omega)$.

## 10.6 Results

In this section, simulation results demonstrating the accuracy of the proposed decomposition in (91) and the proposed transient simulation procedure have been presented. For demonstrating accuracy, Agilent's ADS [99], Synopsys's HSPICE [100], frequency-domain solution have been used as references. The ADS engine is based

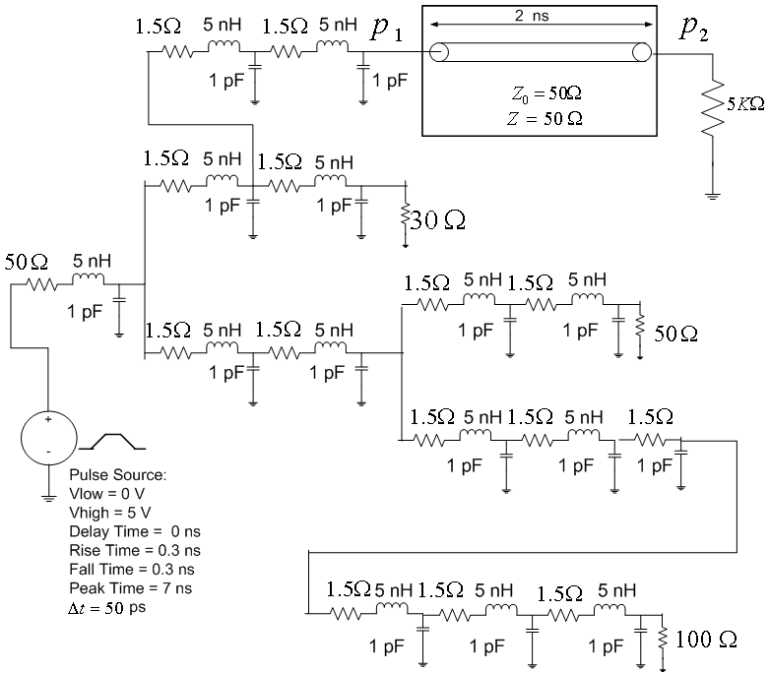


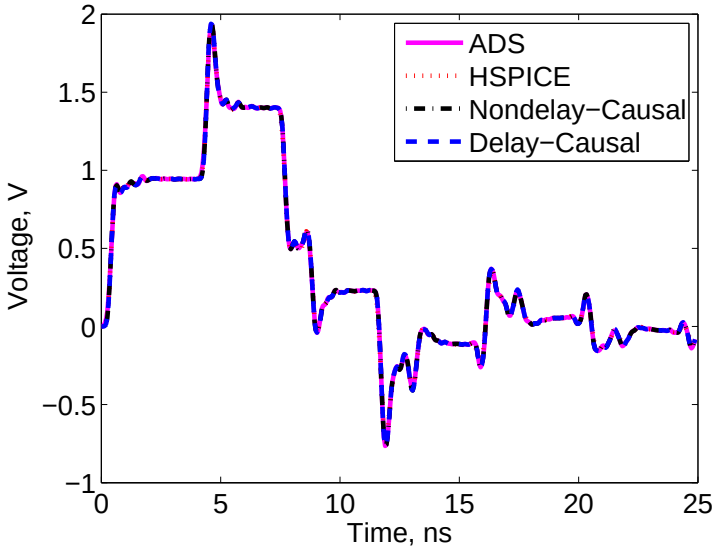

**Figure 88. Test setup for computing pulse response of a lossless transmission line terminated by a distributed RLC circuit. The transmission line is characterized by band-limited two-port causal S-parameters from 0–10 GHz with a frequency step of 1 MHz.**

on the numerical-convolution-based approach. The HSPICE engine (W-element w/ S-parameter input) is based on the recursive-convolution-based approach. Further, in HSPICE simulations, delay is extracted first before rational-function fitting. Group delay is used for this purpose.

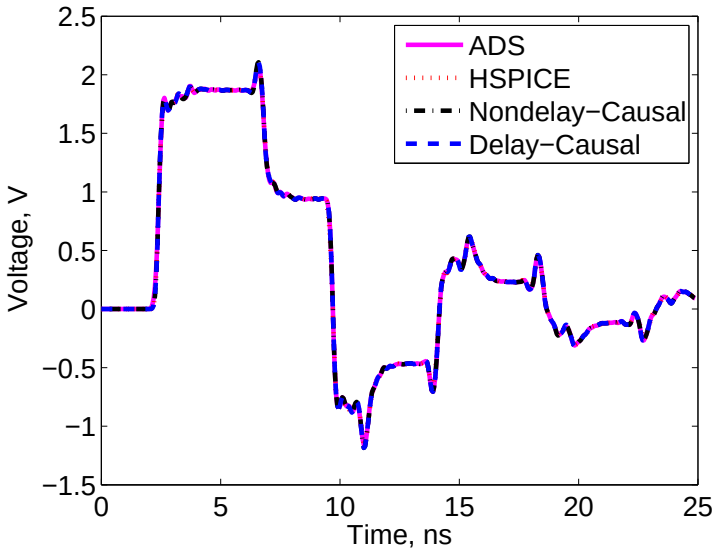
### 10.6.1 Demonstration of Capability to Handle Arbitrary Terminations

First, the accuracy of the proposed method in handling complicated terminations and in extracting the propagation delay is demonstrated. For this demonstration, an example is chosen such that the decomposition in (86) alone is sufficient for the reconstruction of the frequency response. As an example, the pulse response of a lossless transmission line (see Figure 88) is considered. The propagation delay in this line is 2 ns. The average delay extracted using (83) is also 2 ns.

The source termination in Figure 88 is an example of the kind of termination for which it is difficult to use an SFG-based approach, as it is difficult to compute the




(a) Voltage at the near end of the transmission line, i.e., at $p_1$ in Figure 88.




(b) Voltage at the far end of the transmission line, i.e., at $p_2$ in Figure 88.

**Figure 89. Comparison of pulse responses at $p_1$ and $p_2$ in Figure 88 between the proposed method ('Delay-Causal') and ADS, HSPICE, and nondelay-causal simulations.**

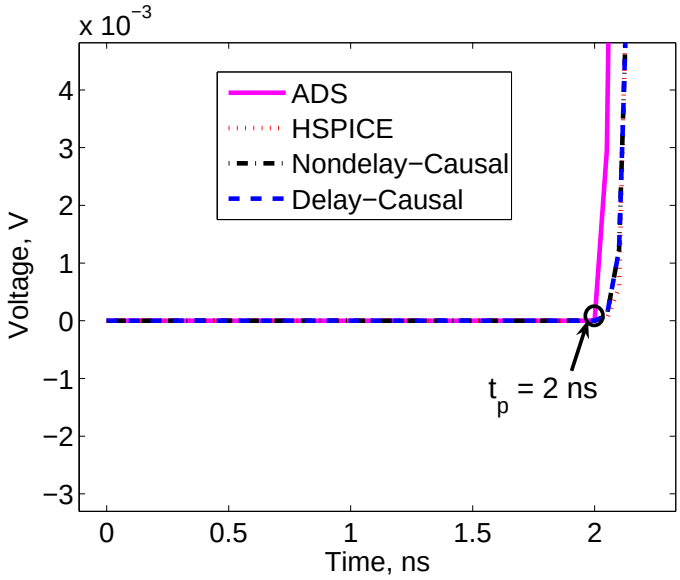
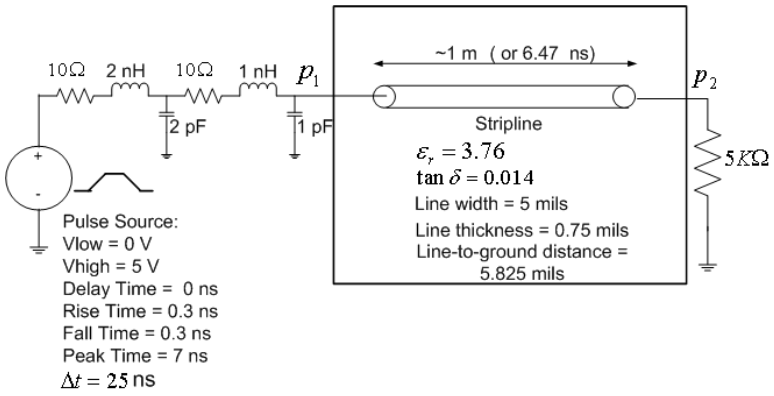



**Figure 90. Zoomed-in voltage at $p_2$ from Figure 89(b) between 0–2 ns. Note the propagation delay of 2 ns through the line is captured in the 'Delay-Causal' results.**

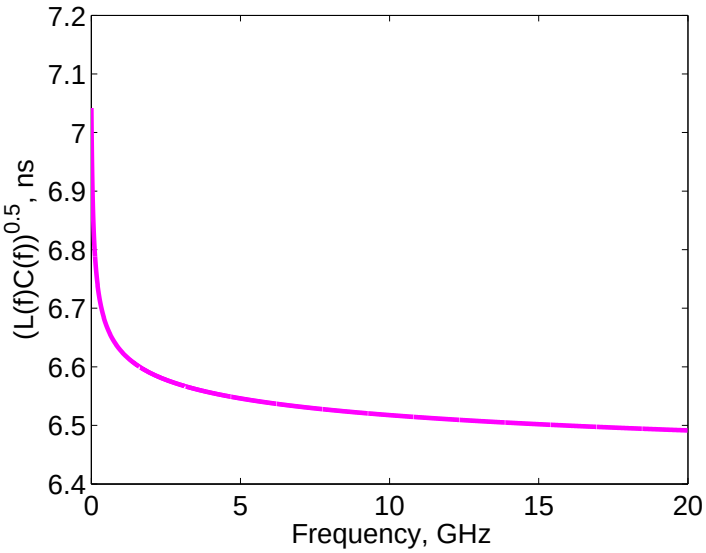
Thevenin's equivalent circuit for the source. On the other hand, in the MNA-based approach, no such difficulty is present. The voltages at both the near end and the far end of the line (ports $p_1$ and $p_2$ in Figure 88) are computed using both delay-causal and nondelay-causal impulse responses. For a lossless transmission line and for an available bandwidth $f_c = k\frac{1}{t_p}$, where $t_p$ is the propagation delay, and $k$ is a positive integer, the nondelay-causal impulse response is automatically delay-causal. That is, in such a situation, no explicit delay extraction and enforcement are needed. Therefore, for this example ($f_c = 20\frac{1}{t_p}$), the nondelay-causal results are the most accurate with respect to handling the f.d. data. To compare the accuracy of the whole system, which includes even the terminations, ADS and HSPICE are used. For this example, the ADS results are expected to be delay-causal for the same reason mentioned above. Therefore, both ADS and HSPICE results can be reliable reference solutions. The delay-causal (denoted as 'Delay-Causal') and nondelay-causal voltages ('Nondelay-Causal') are compared with those obtained from ADS ('ADS') and from HSPICE ('HSPICE') in Figures 89 and 90. From Figures 89(a-b), it can be observed that the 'Delay-Causal' results from the proposed procedure match closely with the

other results. From Figure 90, it can be observed that propagation delay (= 2 ns) is captured exactly in the 'Delay-Causal' results. This example demonstrates the accuracy of the proposed formulation in handling complicated terminations and in extracting (and enforcing) the propagation delay.

Next, the accuracy of the proposed transient simulation has been demonstrated for a dispersive transmission line characterized by causal data. As an example, the pulse response of a lossy strip line (see Figure 91(a)) is considered. The product of the frequency-dependent inductance and capacitance is shown in Figure 91(b). A lossless stripline of the same length would have a delay of approximately 6.47 ns. The average delay extracted using (83) is 6.5 ns. The voltages at both the near end and the far end of the line are computed using both delay-causal and nondelay-causal impulse responses. These voltages are compared with those obtained from ADS and HSPICE in Figures 92 and 93. Unlike the previous example, the transmission line is lossy. Therefore, it is not possible to exploit periodicity to get a reference solution. Therefore, the 'Nondelay-Causal' results may not be reliable as a reference solution with respect to handling the f.d. data. Therefore, the problem was also solved in the frequency domain, and the frequency-domain voltages are converted to the corresponding time domain results using the IFFT. Since the IFFT results inherently denote a circular convolution, care has been taken to make these results perform a linear convolution (see [89], pp. 580, Figure 8.18). The IFFT results are used as the reference. From Figures 92(a-b), it can be observed that the 'Delay-Causal' and 'Nondelay-Causal' results from the proposed procedure match closely with the results from both ADS and IFFT. However, the propagation delay (= 6.5 ns) is captured in the 'Delay-Causal' (Proposed) results (see Figure 93) but not in the 'Nondelay-Causal' and 'ADS' results. It is to noted that the nondelay-causal formulation is same as the delay-causal formulation except for the the delay extraction and enforcement in the latter. This comparison shows that unless the propagation delay is extracted and

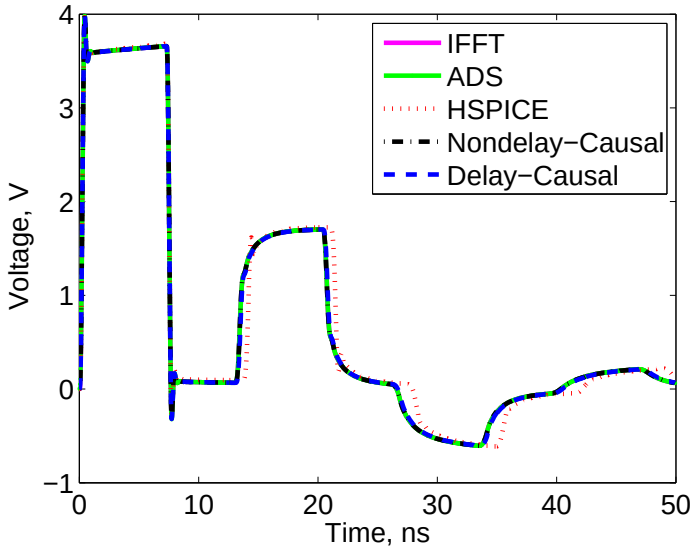



(a) Test setup.

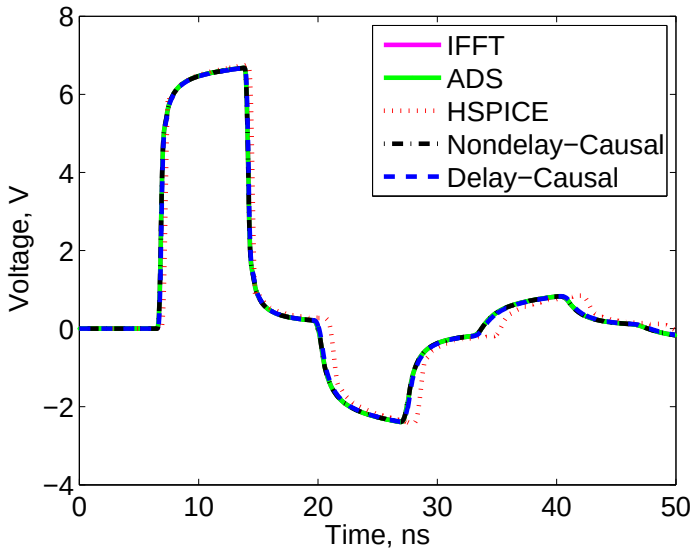



(b) Square root of the product of frequency-dependent inductance and capacitance.

**Figure 91. Test setup of pulse response of a lossy transmission line terminated by a distributed RLC circuit. The transmission line is characterized by band-limited two-port causal S-parameters from 0–20 GHz with a frequency step of 1 MHz.**




(a) Voltage at the near end of the transmission line, i.e., at $p_1$ in Figure 91.




(b) Voltage at the far end of the transmission line, i.e., at $p_2$ in Figure 91.

**Figure 92. Comparison of pulse responses of the set up in Figure 91 between the proposed method ('Delay-Causal') and ADS, HSPICE, frequency-domain solution ('IFFT'), and nondelay-causal simulations.**

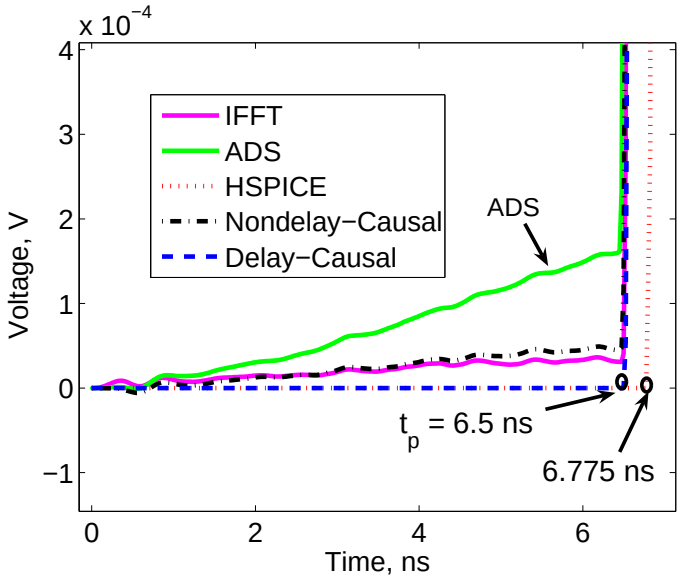
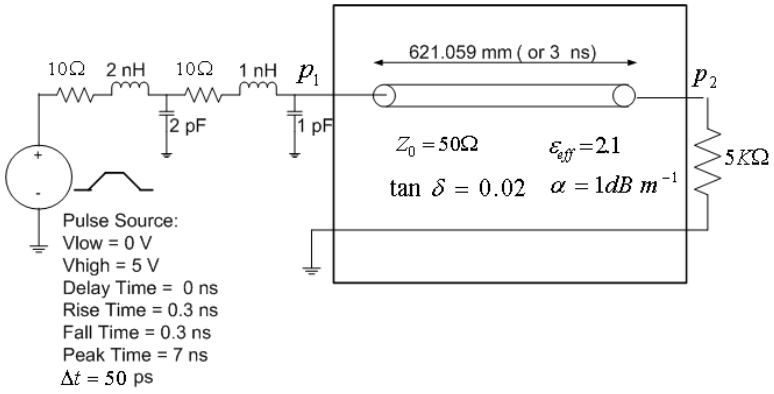



**Figure 93. Zoomed-in voltage at $p_2$ from Figure 92(b) between 0–7 ns. Note the propagation delay of 6.5 ns through the line is captured in the 'Delay-Causal' results only.**

enforced explicitly, it is usually not captured. It is to be noted that the IFFT results will not capture the propagation delay, as these results are equivalent to performing a linear convolution without the delay enforcement. Therefore, the IFFT results would be similar to the nondelay-causal results in terms of delay enforcement and accuracy. This example demonstrates the accuracy of the proposed formulation in handling dispersive causal data.

The 'Delay-Causal' results are also compared with 'HSPICE' results in the same figure (Figures 92 and 93). It is to be noted that the HSPICE results are different from the rest of the results in two ways: 1) From Figure 92(a-b), it can be noticed that the HSPICE results are inaccurate compared to the rest of the results. This inaccuracy is because of the approximation involved in fitting a rational-function-system to delayless frequency responses. On the other hand, 'Delay-Causal' results do not suffer from this inaccuracy, as they are obtained using a numerical-convolution-based approach (which does not curve fit responses). 2) From Figure 93, it can be noticed the propagation delay from HSPICE (6.775 ns) is approximately 0.275 ns ($= 11\Delta t$) more than the predicted delay. From the 'IFFT' and 'Nondelay-Causal' results




(a) Test setup.

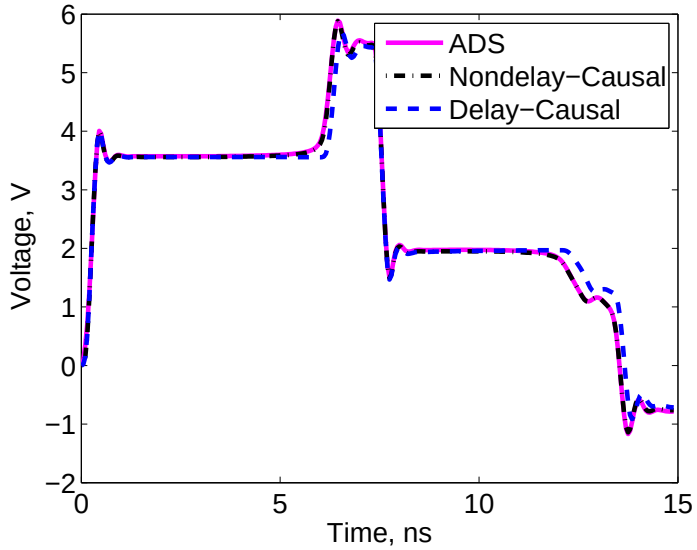
**Figure 94. Pulse response of a lossy transmission line terminated by a distributed RLC circuit. The transmission line is characterized by band-limited two-port noncausal S-parameters from 0–10 GHz with a frequency step of 1 MHz.**

in Figure 93, it can be inferred that the propagation delay of more than 6.5 ns may not be an accurate solution. Therefore, the proposed method can be more accurate than recursive-convolution-based approaches like the W-element in HSPICE in some situations.

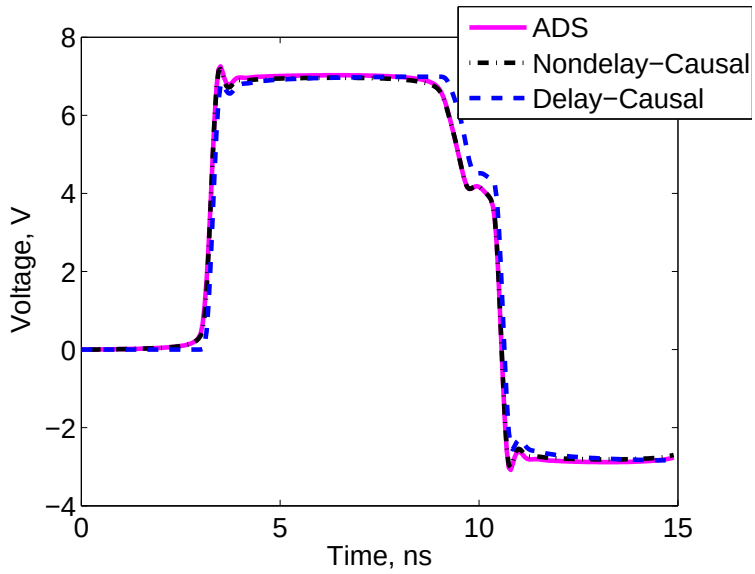
In the previous example, the amplitudes of the nondelay-causal responses before the propagation delay were small (see 'Nondelay-Causal', 'ADS', and 'IFFT' results in Figure 92 for $t < 6.5$ ns). This has to do with the causal nature of the data. However, if the S-parameters were noncausal, this amplitude can be significant [59]. Though the use of noncausal data is not advised [59], there are situations when the user is not aware of the causality of the data. For example, some existing transmission line models (see TLINP model in ADS) inherently produce noncausal data. The proposed technique can be used to get an approximate causal time-domain response given even some noncausal data. This feature is demonstrated in the next example. It is to be noted that Hilbert-transform-based techniques have been employed in the past to handle noncausal data (see [87], [101], [102], [103] [104], [105]). The previous example is repeated for a new transmission line specification (see Figure 94). The noncausal data were obtained by a noncausal circuit model that considers variation

in $R(f)$ and $G(f)$ but ignores variation in $L(f)$ and $G(f)$ (see the model TLINP in ADS). The actual propagation delay for this line is 3 ns. The average delay extracted using (83) is also 3 ns. As can be observed from Figure 96, the 'Delay-Causal' results capture the propagation delay (= 3 ns), while the other two results do not. Also from Figure 96, it can be observed that the amplitudes of the nondelay-causal voltages ('Nondelay-Causal' and 'ADS') before the propagation delay are not small. The close match between the 'Nondelay-Causal' results and 'ADS' results in Figure 95(a-b) demonstrates the accuracy of the proposed formulation without the delay enforcement. The difference observed between the 'ADS' and the 'Delay-Causal' results is because of using noncausal data without delay-causality enforcement in the former. The 'Delay-Causal' results in Figures 95 and 96 are also compared with those from HSPICE in Figures 97 and 98, respectively. From Figure 97(a-b), it can be observed that the 'Delay-Causal' results match closely with the 'HSPICE' results. Also, from Figure 98, it can be noticed both the propagation delay is captured in both the results. The reason for this close agreement with HSPICE is attributed to the delay extraction done prior to rational-function fitting in HSPICE [100]. Such a processing in HSPICE is similar to the one followed in the method-of-characterisics approach [59]. This approach has been demonstrated to yield delay-causal responses for lossy transmission lines in [59].

Until now, the decomposition in (86) is sufficient for all the examples. To demonstrate the need for and the accuracy of the proposed decomposition in (91), a coupled-line example is considered. As an example, the step response in two coupled transmission lines is simulated. Consider the symmetric lossless coupled microstrip transmission lines shown in Figure 99. The geometry of the lines and of the substrate are specified in Figure 99. The coupled lines are modeled by four-port S-parameters from 0 to 4 GHz (Agilent's ADS was used for obtaining the S-parameters). At port $p_1$, a step voltage source is applied. All the ports are terminated with $Z_0$ (= 22 $\Omega$). It was




(a) Voltage at the near end of the transmission line, i.e., at $p_1$ in Figure 94.




(b) Voltage at the far end of the transmission line, i.e., at $p_2$ in Figure 94.

**Figure 95. Comparison of pulse responses of the set up in Figure 94 between the proposed method ('Delay-Causal') and ADS and nondelay-causal simulation.**

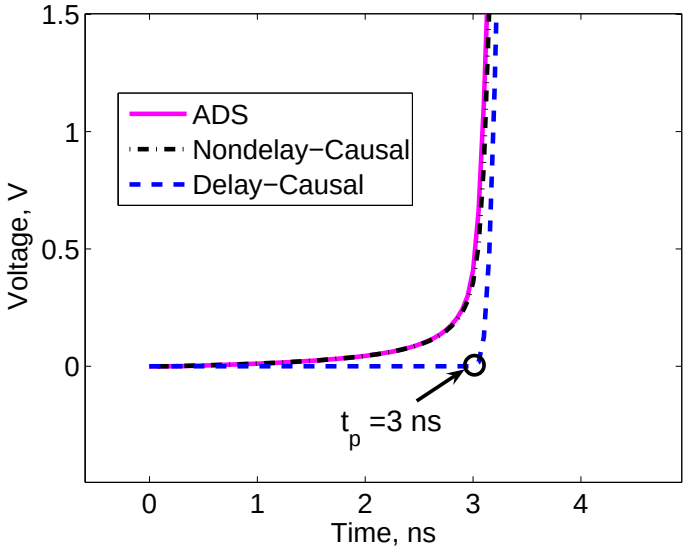
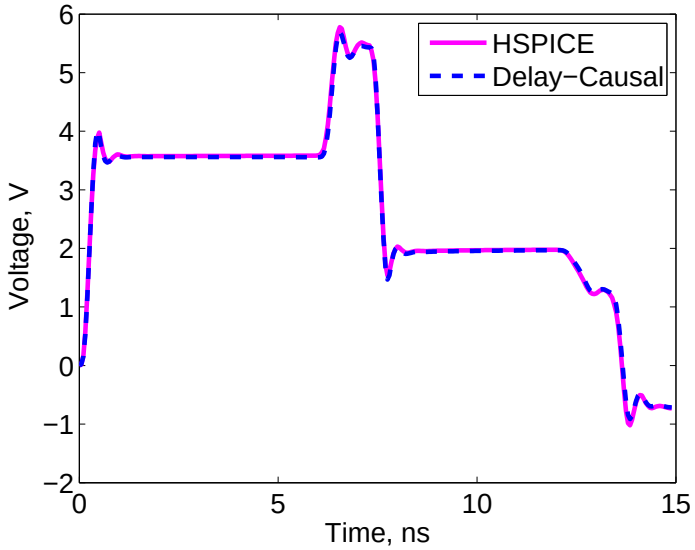


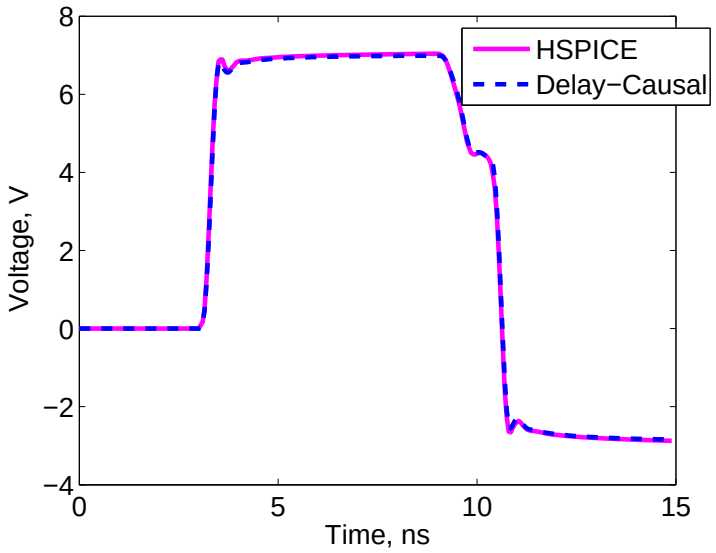

**Figure 96. Zoomed-in voltage at $p_2$ from Figure 95(b) between 0–4 ns. Note the propagation delay of 3 ns through the line is captured in the 'Delay-Causal' results only.**

found that the decomposition in (91) was needed only for the transfer response $S_{41}$. This need would be felt if the constant phase $\theta$ in (91) is shown to take a nonzero value (if $\theta = 0$, the proposed decomposition in (91) is same as the decomposition in (86), hence there is no need for the proposed decomposition). Therefore, the phase $\theta$ is computed for $S_{14}$ $(= S_{41})$, using the numerical procedure described in Section 10.3.3.

The numerical phase extraction procedure in Section 10.3.3 requires computing the angles of the original frequency response and of the minimum-phase response near $\omega \approx 0$ (see (92) specifically). This procedure is described next. In Figure 100 and Figure 101, the procedures to compute the angle of the $S_{14}$ and of its minimum-phase component near $\omega = 0$, respectively, are described. Figure 100 consists of two parts: 1) In the first part, the angle of $S_{14}(\omega)$ is shown for frequencies up to 4 GHz. 2) In the second part, the angle of $S_{14}(\omega)$ is shown only for frequencies near $f = 0$. (see only the results from 'ADS' for the current discussion; the discussion on the 'Analytical' results are deferred for now). From the second part of Figure 100, it can be observed that $\arg[S_{14}(f)] \rightarrow -\frac{\pi}{2}$ as $f \rightarrow 0$.




(a) Voltage at the near end of the transmission line, i.e., at $p_1$ in Figure 94.




(b) Voltage at the far end of the transmission line, i.e., at $p_2$ in Figure 94.

**Figure 97. Comparison of pulse responses of the set up in Figure 94 between the proposed method ('Delay-Causal') and HSPICE.**

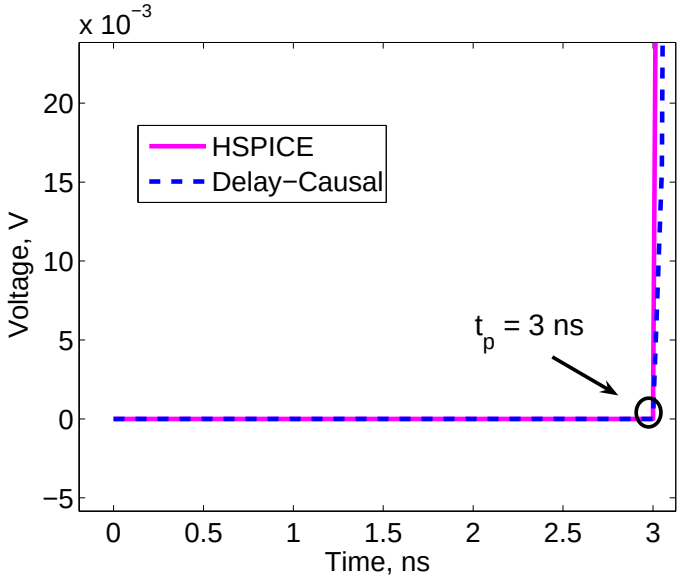



**Figure 98. Zoomed-in voltage at $p_2$ from Figure 97(b) between 0–4 ns.**

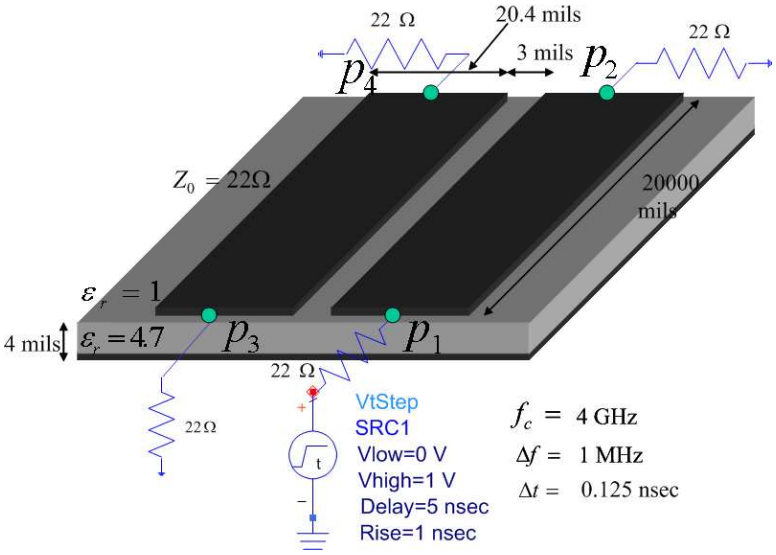



**Figure 99. Test set up of a coupled microstrip transmission line circuit in which the lines are characterized by four-port S-parameters. The symbol $p_i$ refers to port $i$. The circuit is excited by a step source at $p_1$, and the transient voltages at $p_2$ and $p_4$ are computed.**

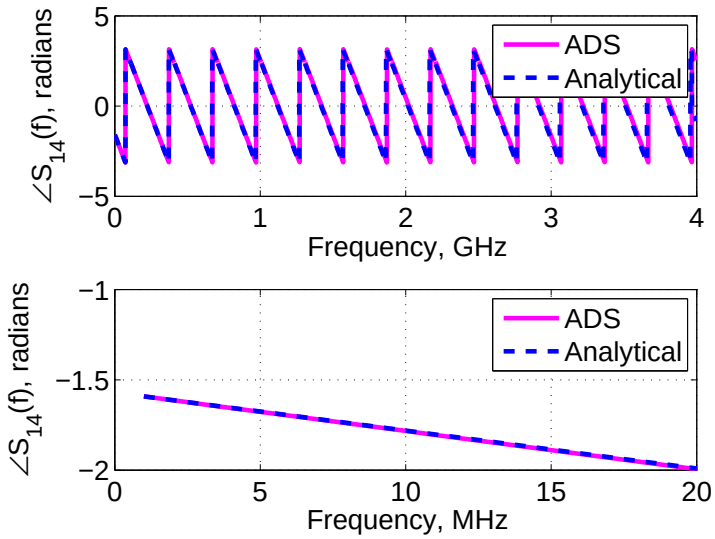
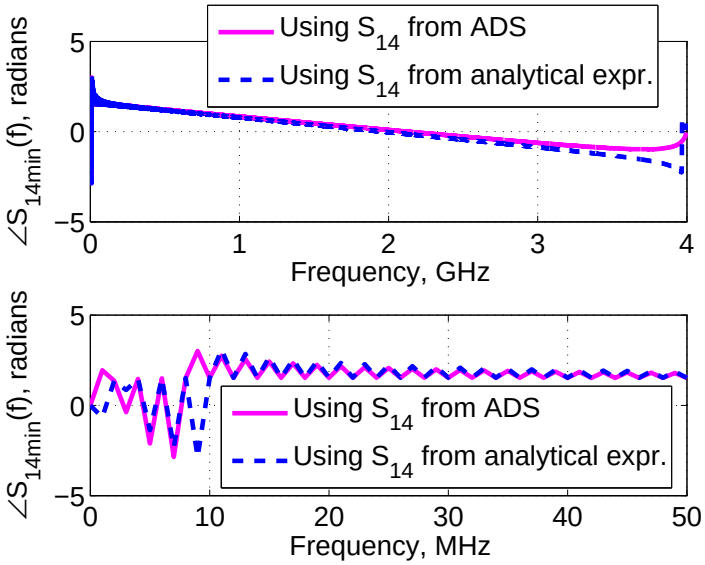


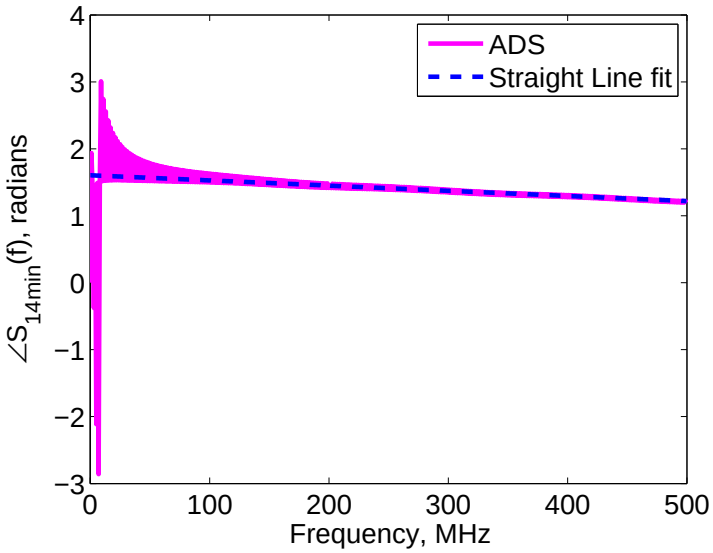

**Figure 100.** $\arg[S_{14}(f)] \rightarrow -\frac{\pi}{2}$ **as** $f \rightarrow 0$**.**

In Figure 101, the procedure to compute the phase of $S_{14\text{min}}(\omega)$ near $\omega = 0$ is described. Once again, in Figure 101, the discussion on '$\cdots$ analytical expr.' is deferred for now. Unlike $\arg[S_{14}(f)]$, $\arg[S_{14\text{min}}(f)]$ is not smooth near $f = 0$ (see both parts of Figure 101). The argument of $S_{14\text{min}}(f)$ is found to have numerical oscillations (triangular oscillations with period $2\Delta f$, see Figure 101) because only a numerical Hilbert transform is being applied and the phase of minimum-phase component changes abruptly at $f = 0$. Hence, an asymptotic value of the phase is computed for $S_{14\text{min}}(f)$. It is found that $\arg[S_{14\text{min}}(f)] \rightarrow \frac{\pi}{2}$ asymptotically as $f \rightarrow 0$, as shown in Figure 101(b) by the intercept of the straight line (fitted to the angle of the minimum-phase response) with $f = 0$ axis. Since the angles $\arg[S_{14}(f)]$ and $\arg[S_{14\text{min}}(f)]$ approach different values as $f \rightarrow 0$, it is obvious from (92) that $\theta \neq 0$ for $S_{14}(\omega)$. Using these phases and using (92), the constant phase $\theta$ is computed to be $-\pi$ for $S_{14}(f)$.

This value of $\theta$ for $S_{14}(\omega)$ can be verified theoretically as follows. The discussion deferred thus far on the analytical results in Figure 100 and Figure 101 is now




(a) Angle of $S_{14_{\min}}(f)$ as a function of $f$.




(b) Straight-line fit to calculate the angle of the minimum-phase component $f \to 0$.

**Figure 101.** $\arg[S_{14_{\min}}(f)] \to \frac{\pi}{2}$ **as** $f \to 0$**.**

explained. From [106], $S_{14}(\omega)$ can be analytically expressed as

$$S_{14}(\omega) = -je^{-j\beta_0 l}\sin\left(\left(\beta_e - \beta_o\right)l\right), \tag{115}$$

where $l$ is the length of the line, and $\beta_0$ is the propagation constant of the uncoupled line, and $\beta_e$ and $\beta_o$ are the even- and odd-mode propagation constants of the coupled lines. These propagation constants are defined as $\beta_0 = \frac{\omega\varepsilon_{\text{eff}}^{\frac{1}{2}}}{c}$, $\beta_e = \frac{\omega\varepsilon_e^{\frac{1}{2}}}{c}$, and $\beta_o = \frac{\omega\varepsilon_o^{\frac{1}{2}}}{c}$, where $c$ is the velocity of light in vacuum, $\varepsilon_{\text{eff}}$ is the effective relative dielectric constant of the lines when they are uncoupled, and $\varepsilon_e$ and $\varepsilon_o$ are the even- and odd-mode relative dielectric constants, respectively, of the lines when they are coupled. For the coupled line in Figure 99, $\varepsilon_{\text{eff}}$ can be computed as 3.9056 from [95]; the dielectric constants $\varepsilon_e$ and $\varepsilon_o$ are obtained from ADS and are 4.191 and 3.604, respectively. To validate the expression in (115), the plots in Figure 100 and Figure 101 are obtained using (115) and are plotted together with the previous plots. From (115), because of the term $-j$, it can be clearly seen that $\arg\left[S_{14}(f)\right] \to -\frac{\pi}{2}$ as $f \to 0$. Also, from (115) and (89), $S_{14_{\min}}(f) = -\Im\left[\text{HT}\left\{\ln\left|\sin\left(\left(\beta_e - \beta_o\right)l\right)\right|\right\}\right]$, for which there is no analytical solution. It can be numerically shown that the argument of the minimum-phase response of a sine function such as $\sin(a\omega)$, where $a$ is independent of $\omega$, approaches $\frac{\pi}{2}$ asymptotically as $f \to 0$. Therefore, the constant phase $\theta = -\pi$ for $S_{14}(\omega)$.

With the need for the decomposition in (91) already established, the effect of using (or not using) the proposed decomposition on the transient results is shown next. The objectives of the following are two fold: 1) to further demonstrate the need for the proposed decomposition using transient results, and 2) to demonstrate the accuracy of the proposed transient simulation procedure with the proposed decomposition. The effect on the transient results can be considerable sometimes, as will be shown in the current example, and can be not-so considerable yet important sometimes, as will be shown in the next example. For this purpose, the voltage at $p_4$ (see Figure 99) is computed. In Figure 102, the voltages at $p_4$ obtained from the decomposition in (86) ("Delay-Causal, LP" in Figure 102) and from the decomposition

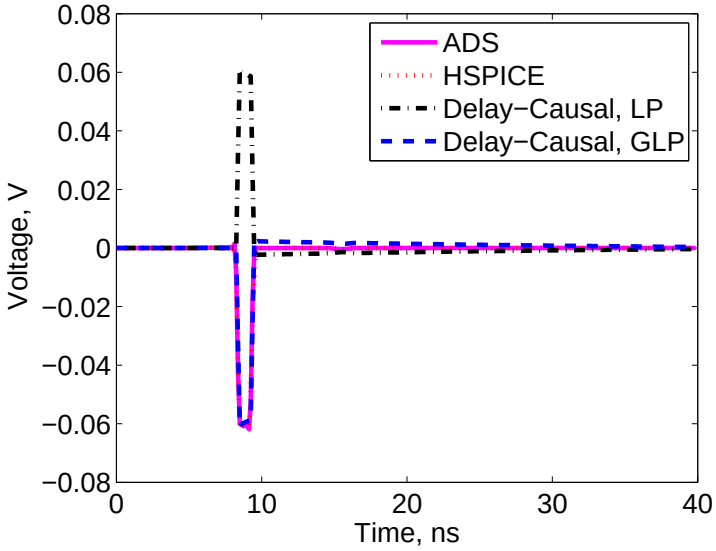
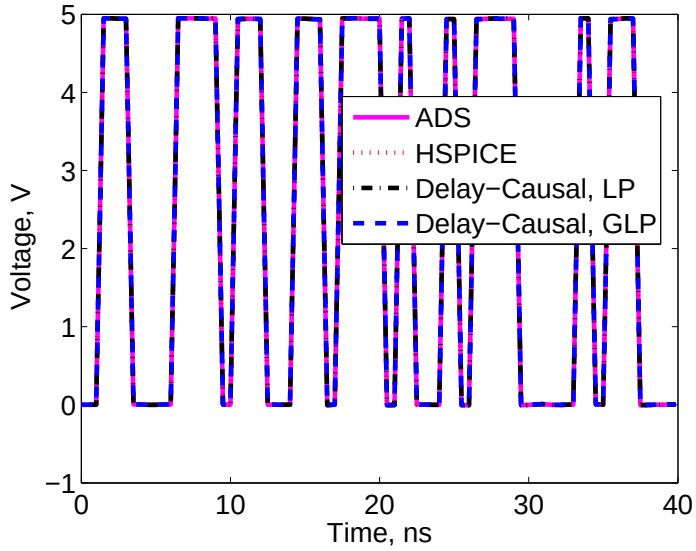
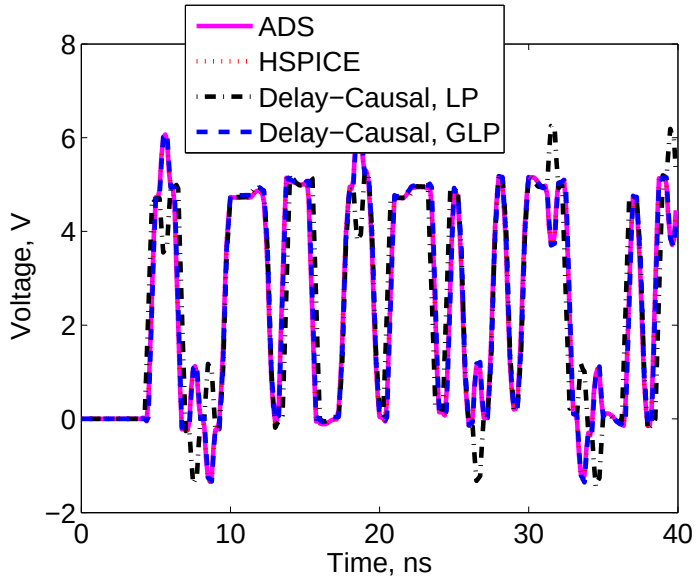



**Figure 102. Comparison of transient results at $p_4$ in Figure 99 using the decompositions in (86) ('Delay-Causal, LP') and (91) ('Delay-Causal, GLP').**

in (91) ("Delay-Causal, GLP" in Figure 102) are compared. The voltages from ADS and from HSPICE are used as a reference. From Figure 102, it can be observed that the voltage in the case "Delay-Causal, LP" differs in sign from the voltage in the case "Delay-Causal, GLP". Thus, not using the proposed decomposition can result in an incorrect transient result. In the previous example, the correct result among the results obtained with and without the decomposition could be found with the knowledge of the far-end crosstalk because of a step source. In the next example, a case is shown where such finding would be hard. As an example, the coupled lines in Figure 99 are excited at $p_1$ and $p_3$ by pseudorandom bit sources, and the voltages at all the ports are computed. Each source has a series resistance of 0.25 Ω, an amplitude of 5 V, and a rise and fall time of 0.5 ns each. The time step of the simulation is same as in the previous example. In Figures 103 and 104, the voltages at all the four ports are compared with those from ADS and HSPICE. From Figure 103(b) and Figure 104(b), it can be seen that that the results from the case 'Delay-Causal, GLP' (dash) match with those from ADS (solid) and HSPICE (dot), demonstrating the accuracy of the transient results with the proposed decomposition. It can be noticed that the

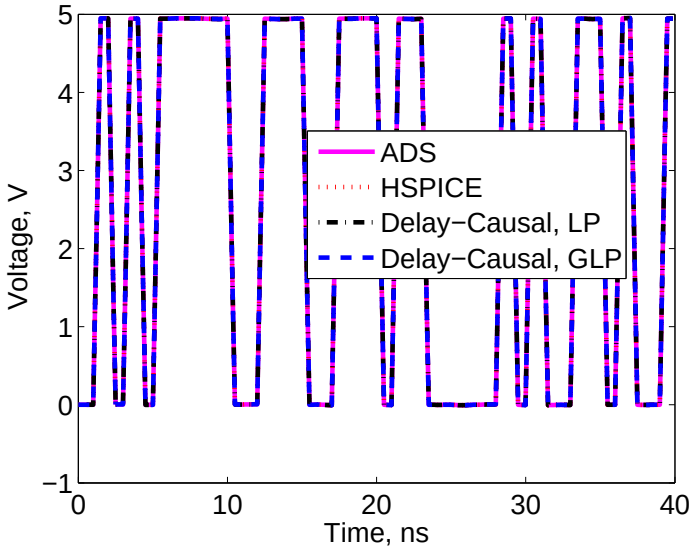



(a) Voltage at $p_1$.

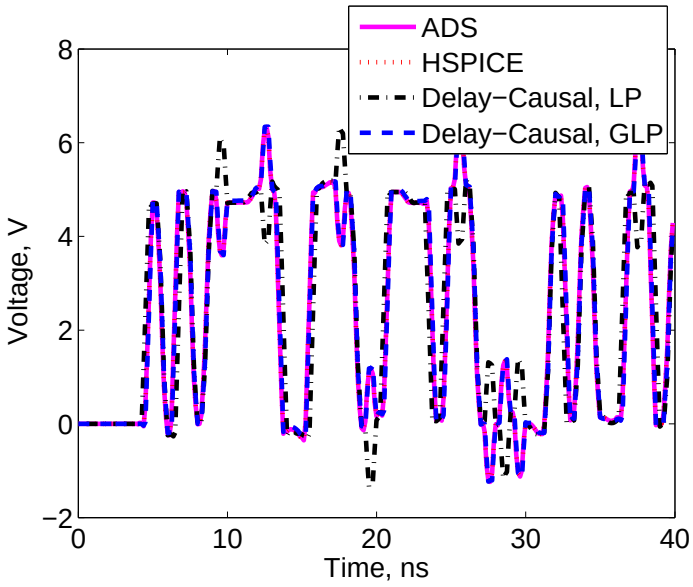



(b) Voltage at $p_2$.

**Figure 103. Comparison of transient responses at ports $p_1$ and $p_2$ obtained with linear-phase condition and with generalized linear-phase condition. Example is a coupled transmission line excited by pseudorandom bit patterns.**




(a) Voltage at $p_3$.




(b) Voltage at $p_4$.

**Figure 104. Comparison of transient responses at ports $p_3$ and $p_4$obtained with linear-phase condition and with generalized linear-phase condition. Example is a coupled transmission line excited by pseudorandom bit patterns.**

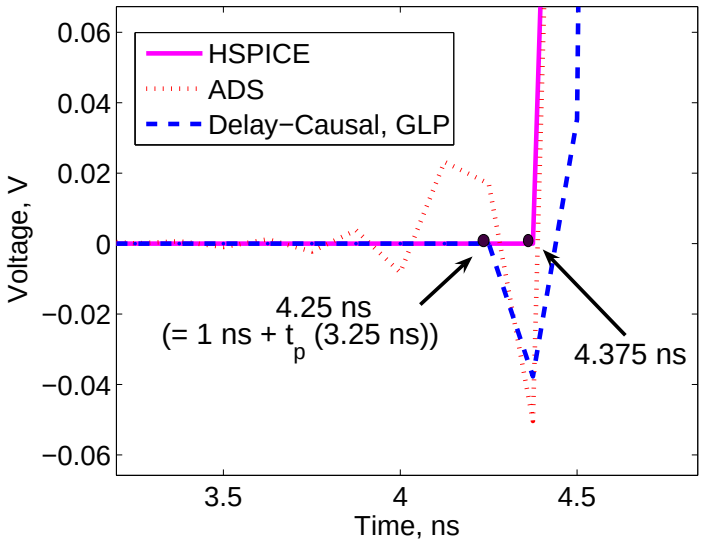



**Figure 105. Comparison of voltage at $p_4$ between 3.5–4.5ns from different methods. Approximate propagation delay is captured in the delay-causal result.**

voltage at $p_4$ ($p_3$) from the case 'Delay-Causal, LP' (dash-dot) have opposite voltage excursions compared to the voltages from the other cases. The difference between the cases 'Delay-Causal, LP' and 'Delay-Causal, GLP' is not as considerable as it was in the previous example, yet may be important. Moreover, this difference can become considerable if the rise time of the voltage source is reduced, as the crosstalk is inversely proportional to rise time in a two-coupled transmission line [107]. Therefore, the proposed decomposition in (91) is necessary for accurate delay-causal transient simulation in examples such as the coupled transmission lines.

Though the 'Delay-Causal' results match well with ADS (Figures 103 and 104), propagation delay is captured only in the former. To demonstrate this difference, the voltage at $p_4$ is compared between the 'Delay-Causal' and 'ADS' results for 3.5–4.5ns in Figure 105. The propagation delays computed with $\epsilon_{\text{eff}}$, $\epsilon_e$, and $\epsilon_o$ are 3.349 s, 3.469 s, and 3.217 s, respectively. The average delay computed from (83) is 3.2713 s for $S_{41}(f)$. The sources at $p_1$ and $p_3$ are nonzero only after 1 ns. Therefore, the voltage at $p_4$ should be nonzero only after an additional time delay equal to the propagation delay. From Figure 105, it can be observed that the propagation delay (approximated

to the nearest multiple of $\Delta t$ in Figure 105) is captured in the 'Delay-Causal' result but not in the 'ADS' result. The HSPICE result has an approximate propagation delay of 3.375 ns (see Figure 105).

The proposed method, unlike the W-element in HSPICE, can be theoretically applied to even nontransmission line examples.

## 10.7 Summary

A numerical-convolution-based procedure has been proposed for the accurate transient simulation of interconnects characterized by b.l.f.d. data and terminated by arbitrary equivalent circuits. Propagation delay is enforced in the transient results by obtaining a causal impulse response through a new minimum-phase/all-pass decomposition of the frequency data, extracting the delay from the data, and enforcing the delay in the causal impulse response. In this decomposition, a new form for the all-pass component has been proposed that preserves the sign of the original frequency response in the reconstructed response, unlike the prior approaches, leading to an accurate transient result. This new form is shown to be essential in computing the far-end crosstalk in coupled microstrip transmission lines. Arbitrary terminations are conveniently handled by integrating the numerical convolution in a modified nodal analysis (MNA) framework, a framework used by commercial circuit simulators, through a new transient simulation formulation. Numerical results demonstrating the improved accuracy and capability of the proposed procedure compared to the prior approach and to the commercial circuit simulators Agilent's advanced design software and Synopsys's HSPICE have been shown.

# CHAPTER 11

# CAUSALITY ENFORCEMENT FOR SELF RESPONSES

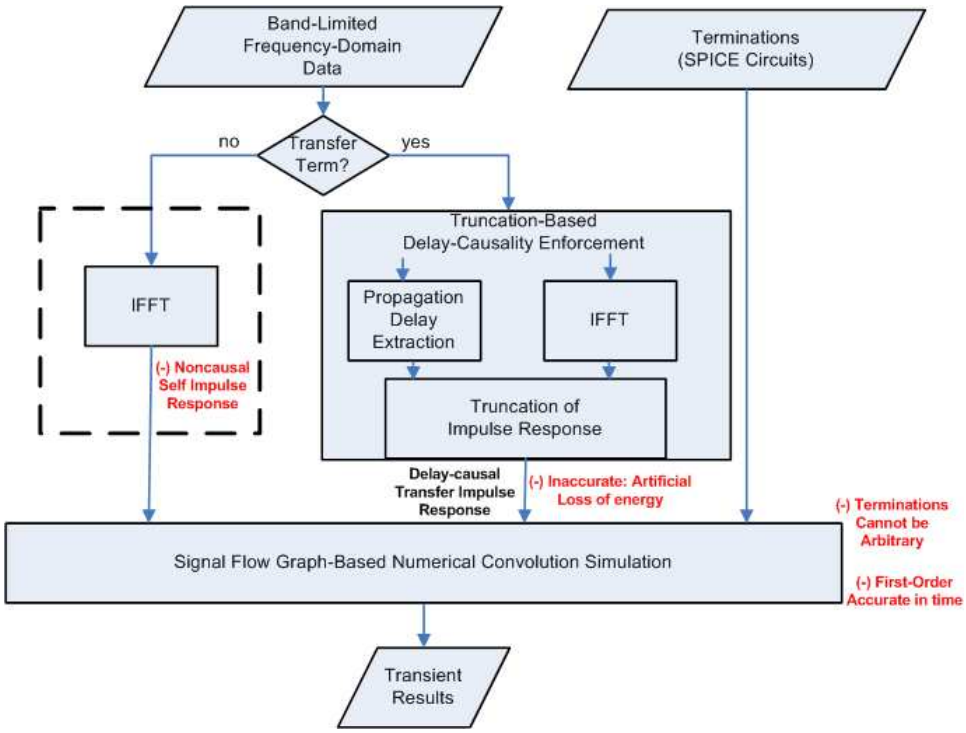
## 11.1 Introduction

In Chapters 9 and 10, the delay causality violations induced by finite bandwidth and frequency-domain windowing on only transfer responses were addressed, and *the causality violations on self responses were not addressed*. The reason for ignoring self responses thus far (in Chapters 9 and 10 and also in [55], [56], [86], [57]) is that the delay extraction and enforcement are not needed for them, as each of these responses has zero propagation delay. *Implicit in this treatment of self responses is that they are assumed to be causal in spite of the factors that induce causality violations acting upon them.*

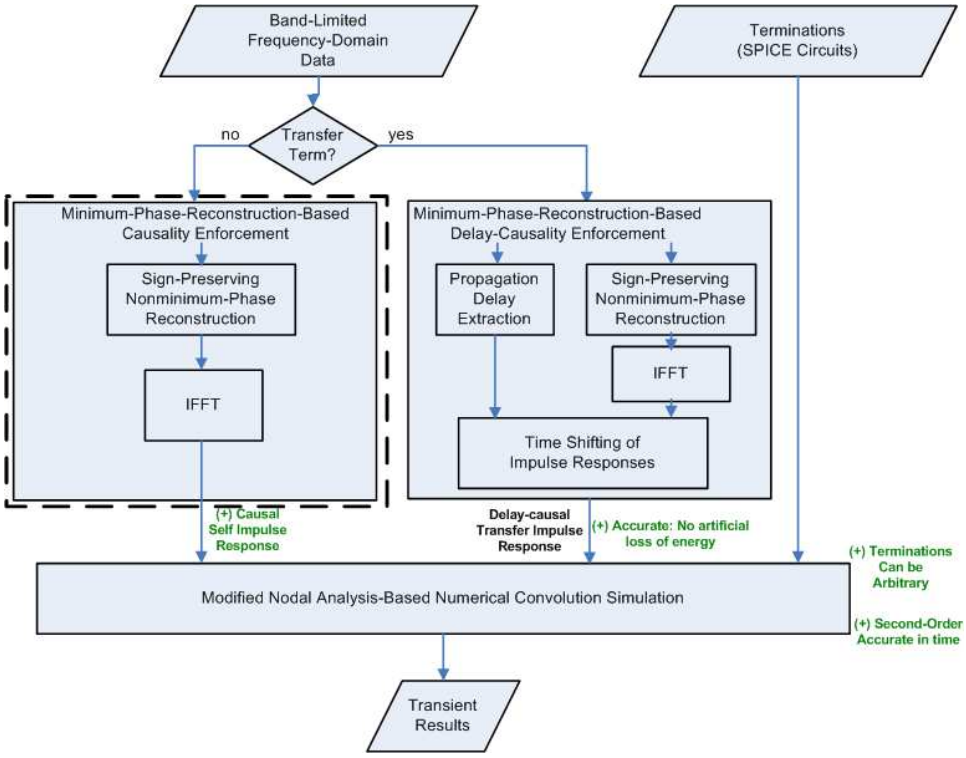
In this chapter, the causality of self responses is studied. It has been shown, contrary to the prior-held belief, that self impulse responses can become noncausal because of the same factors that induce causality violations in transfer responses. Therefore, self frequency responses should also be treated like the way transfer responses are. In this chapter, the sign-preserving nonminimum-phase reconstruction (NMPR) technique, proposed in Chapter 10 for transfer responses, has been proposed for enforcing causality in self responses. Numerical results demonstrating the accuracy of the proposed technique have been presented. *The contribution of this chapter is the causality enforcement of self responses.* The focus of this chapter is also described in Figure 106.

## 11.2 Background

For many interconnects, their frequency responses (FRs) are available as tabulated multiport f.d. data for a limited set of frequencies. Transient simulation with such data is preferable for signal integrity (SI) analysis of interconnects. Capturing the




(a) Prior approach [55], [56], [57].




(b) Proposed approach in this dissertation.

**Figure 106. Comparison of the prior and proposed approach in numerical-convolution-based causal transient simulation of band-limited data. The focus of this chapter is the region marked within the dashed rectangle.**

propagation delay (PD) through interconnects in this simulation is important. The PDs between ports are captured in the simulation if port-to-port impulse responses (IRs) are delay-causal (i.e., zero for times less than the propagation delay). However, IRs computed through the inverse discrete fourier transform (IDFT) of the f.d. data are usually not delay-causal if the f.d. data are noncausal or are band limited [55], [57], [86], [108]. As the data are mostly band limited, delay-causality (DCL) has to be enforced.

Delay-causality enforcement involves two steps: 1) Making transfer impulse responses delay-causal and 2) making self impulse responses causal (a special case of delay-causal with zero PD). In [55], [57], [86], [108], DCL enforcement techniques for this simulation have been proposed for handling delay-causality violations (DCLVs) because of causal but b.l. data. All these approaches enforce delay-causality of transfer impulse responses explicitly. Such an enforcement requires extracting propagation delays from transfer FRs and using them to produce delay-causal transfer IRs. However, all these approaches assume causality of self impulse responses and hence do not enforce it explicitly. This assumption is found to be true only for some situations.

Self IRs from causal scattering parameter (S-parameter) data have been found to be noncausal either when f.d. windowing is applied or when the bandwidth (BW) of the data is not sufficient. Frequency-domain data are subjected to a f.d. windowing to obtain a smooth transient response and sometimes to obtain a stable transient response [109]. When self IRs are not causal, transient results are not physical and hence are not desirable. Therefore, the causality (CL) of self IRs has to be enforced.

Causality of self IRs are implicitly enforced in [55], [86], [57], [108] because of the truncation of the noncausal part of IRs ($\hat{h}(t)$) implicit in the causal convolution integral $\left(\int_{\tau=0}^{t} \hat{h}(\tau) x(t-\tau)\, d\tau\right)$. However, t.d. truncation may not preserve the energy in the original FR and can cause a change in the DC level ($\propto$ area under the IR) of transient (step) responses. Moving the noncausal part of the IR into its causal part is

an option to preserve energy, but the effect of this option on accuracy is not known.

Applying the procedures originally adopted in [55], [57] (for transfer responses) to self responses can make transient results inaccurate: In [55], transfer IRs before the port-to-port PD are truncated. Therefore, this approach would result in an inaccuracy similar to the one described above. In [57], a nonminimum-phase reconstruction (NMPR) technique has been employed. In this technique, a causal IR is first obtained from the minimum-phase reconstruction (MPR) of the FR [87] and is later shifted by the PD. However, for self S-parameter-based FRs, this technique does not capture any leading negative sign in FRs, making the resulting transient results inaccurate.

In [108] (also in Chapter 10), a constant phase term with unity magnitude has been added to the NMPR technique to account for negative signs in transfer responses. In this chapter, the sign-preserving NMPR (SP-NMPR) technique proposed in [108] for transfer responses has been employed to enforce causality of self IRs. The proposed technique has been shown to handle DCLVs because of b.l. data and f.d. windowing.

*The contributions of this chapter are the following:*

1. *Establishing that frequency-domain windowing can cause and worsen causality violations (CLVs) in transient simulation.*

2. *Accurate delay-causal transient simulation using the sign-preserving NMPR technique even for self responses (unlike [108]).*

The rest of this chapter is organized as follows: In Section 11.3, the f.d. windowing-induced CLV is explained. In Section 11.4, numerical results demonstrating the CLV caused by f.d. windowing and the accuracy of the proposed technique have been presented. In Section 11.5, the conclusions of this chapter have been reported.

## 11.3 Delay-Causality Violations

Consider a linear time-invariant system with propagation delay $t_p$ whose frequency response, $H(f)$, is known at uniformly-spaced frequencies between zero and a maximum frequency, $f_c$. Let $h(t)$ be the IDFT of $H(f)$. The response $h(t)$ is delay-causal for a physical system. Let $\breve{H}(f)$ denote $H(f)$ when the latter is not causal. Let $G(f)$ and $W(f)$ denote gate function and window function [89], respectively, with a cut-off frequency $f_c$. Then the frequency response before IDFT, $\hat{H}(f)$, can be expressed as

$$\hat{H}(f) = \breve{H}(f)\, G(f)\, W(f)\,. \tag{116}$$

Let $\breve{h}(t)$, $g(t)$, and $w(t)$ be the IDFTs of the $\breve{H}(f)$, $G(f)$, and $W(f)$, respectively. Then, the IDFT of (116), $\hat{h}(t)$, can be written as

$$\hat{h}(t) = \breve{h}(t) * g(t) * w(t)\,, \tag{117}$$

where the symbol $'*'$ denotes the linear convolution operator. The response $\hat{h}(t)$ in (117) is delay-causal (or causal if $t_p = 0$) only if $\breve{h}(t)$ is delay-causal (or causal if $t_p = 0$), and $g(t)$ and $w(t)$ are causal. The former condition is true when $\breve{H}(f) = H(f)$ and the latter when $f_c = \infty$. The latter condition arises because of the following reason: Both $g(t)$ (a sinc function) and $w(t)$ are real and even functions in time and therefore are noncausal, as their fourier transforms are real (and positive) and even functions in frequency [89]. They become an impulse function centered at $t = 0$, a causal function, for $f_c = \infty$. Therefore, with band-limited data (i.e., $f_c \neq \infty$), $\hat{h}(t)$ may not be delay-causal even if $\breve{h}(t)$ is delay-causal. Under some cases, even with a $f_c \neq \infty$, $\hat{h}(t)$ can be delay-causal provided $W(f)$ is not present. Such a case happens for an ideal transmission line with $f_c = \frac{1}{t_p}$ (transfer $t_p$) with $\Delta t = \frac{1}{2f_c}$ and $t_p = m\Delta t$, where $\Delta t$ is the time step, and $m$ and $k$ are positive integers (see Figure 109(a)). However, when $W(f)$ is also present, $\hat{h}(t)$ can become nondelay-causal (see Figure 111(a)). This also means that the self $\hat{h}(t)$ can become noncausal for the above case. Unlike $G(f)$, the $W(f)$-induced CLVs can worsen when $W(f)$ is made stronger. It is

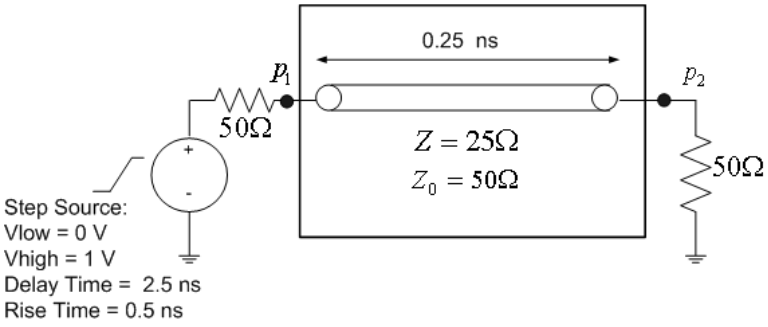
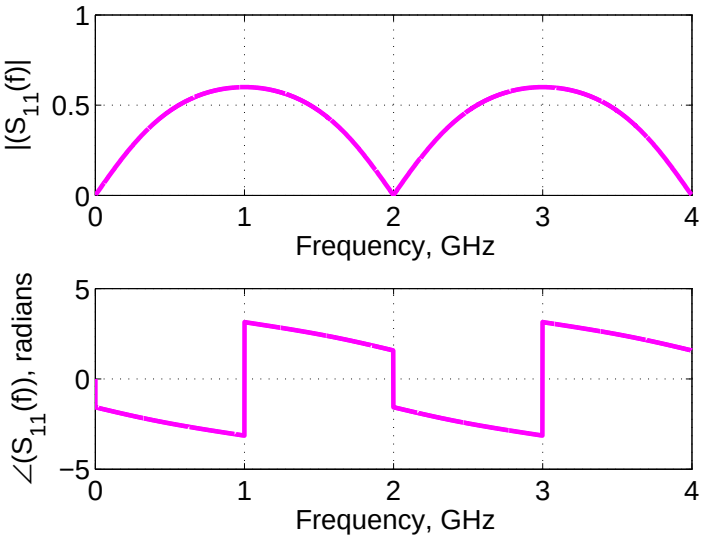



**Figure 107. Test setup: Step response of a lossless transmission line (2-port data) with $t_p = 0.25$ ns.**

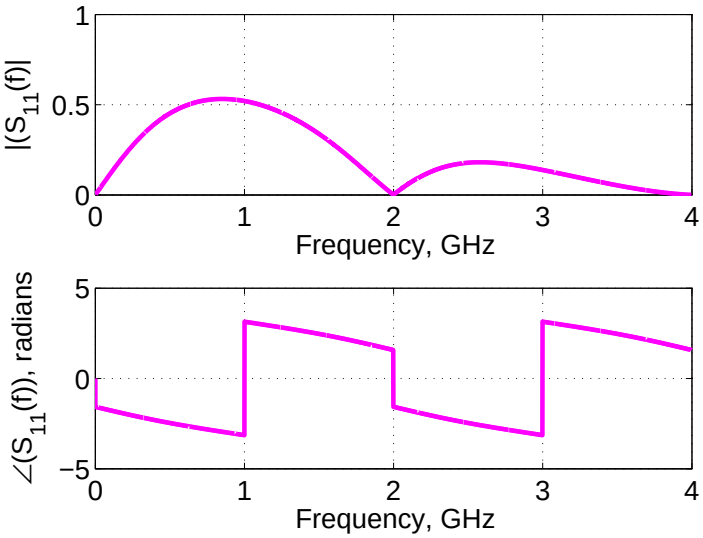
to be noted that $W(f)$, unlike $G(f)$, can make the f.d. data noncausal, as it provides a frequency-dependent attenuation of the amplitude of the FR without altering its phase (see Figure 107(b-c)). Such transformed data may not satisfy Kramer-Kronig relationships and therefore may not be causal.

## 11.4 Results

In this section, numerical results demonstrating 1) the noncausality of self IRs, and 2) the accuracy of the transient simulation using the proposed technique are presented. Towards to this end, the step response of a lossless transmission line (see Figure 107) is computed in the presence of DCLVs as a result of $G(f)$ and $W(f)$. The reflection S-parameter, $S_{11}$, would have a leading negative sign if the characteristic impedance, $Z$, of the line is less than the reference impedance, $Z_0$ [110] (See also the $-\pi$ phase of $S_{11}(f)$ for $f \approx 0$ in Figure 108(a)). When $f_c = 4$ GHz $(= 1/t_p)$ and $W(f)$ is not applied, no DCLVs are observed. To excite $G(f)$-induced DCLV, $f_c$ is set to 3.5 GHz, and $W(f)$ is not applied. To excite $W(f)$-induced DCLV, $W(f)$ is set to a Kaiser window (with a shape parameter, $\beta = 5$ [89]), and $f_c = 4$ GHz. (see Figure 108(b) for $S_{11}$ with this windowing). The transient results are computed using three different techniques: 1) without any CL enforcement for self IRs ('-'), 2) with CL enforcement using NMPR technique ('NMPR'), and 3) with CL enforcement using SP-NMPR technique ('SP-NMPR'). All these techniques enforce DCL for transfer




(a) $S_{11}(f)$ (magnitude in top, phase in bottom), $f_c = 4$ GHz, no $W(f)$.




(b) $S_{11}(f)$, $f_c = 4$ GHz, with $W(f)$. Kaiser window was used.

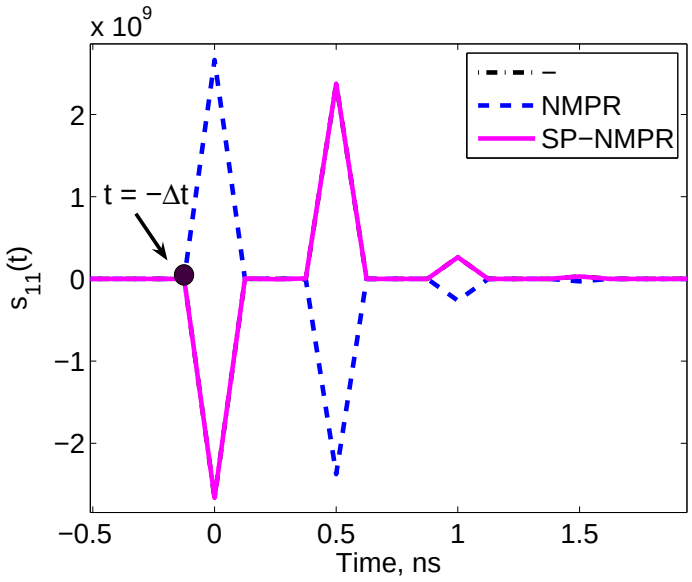
**Figure 108.** $S_{11}(f)$ **with and without** $W(f)$**.**

IRs using the SP-NMPR technique. The time step of the simulation, $\Delta t$, is chosen as $\frac{1}{2f_c}$. The self impulse response, $s_{11}(t)$, and the step response at port $p_1$ are computed for the following three cases: 1) No DCLVs (see Figure 109); 2) $G(f)$-induced DCLV (see Figure 110); and 3) $W(f)$-induced DCLV (see Figure 111).

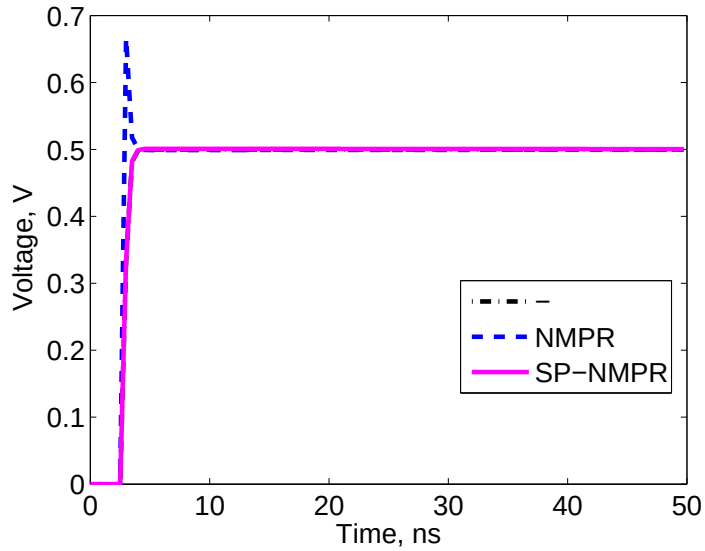
The response $s_{11}(t)$ should ideally consist of periodic impulse streams with period $2t_p$ (see Figure 109(a)). The step responses at both the ports should settle down to 0.5 V after some reflections (ideally after infinite reflections, but practically after a couple of reflections) (see Figure 109(b-c)). From Figure {109-111}, the following can be observed:

1. When CL enforcement is not performed for $s_{11}(t)$, $s_{11}(t)$ can be noncausal either because of $G(f)$ (see Figure 110(a)) or because of $W(f)$ (Figure 111(a)). The step responses in such situation settle to an incorrect final value (see Figure 110(b) and 111(b)) because of the truncation of the noncausal part of $s_{11}(t)$, implicitly enforced by the convolution integral. It is to be noted that $W(f)$ does not affect low-frequency components much and does not affect the zero-frequency component strictly (see Figure 107(b-c)). Therefore, the inaccurate DC level is an artifact of the CL nonenforcement. Since $W(f)$ makes $s_{11}(t)$ more noncausal than $G(f)$, the disparity in the DC values of the step responses with $W(f)$ is more. This disparity increases when $W(f)$ is made stronger. As DCLVs because of $G(f)$ and $W(f)$ cannot be known apriori for generic data, not enforcing causality on self responses can lead to inaccurate results.

2. When CL enforcement is performed for $s_{11}(t)$ but using the NMPR technique without the sign preservation (as in a direct extension from [57]), $s_{11}(t)$ is always causal (see Figure {109-111}(a) for zero response for $t \leq -\Delta t$). However, this $s_{11}(t)$ differs from that of the proposed technique by a negative sign (see Figure {109-111}(a)). This difference manifests itself as a spurious spike in the voltage at $p_1$ (see Figure {109-111}(b)). However, the final values of the step responses

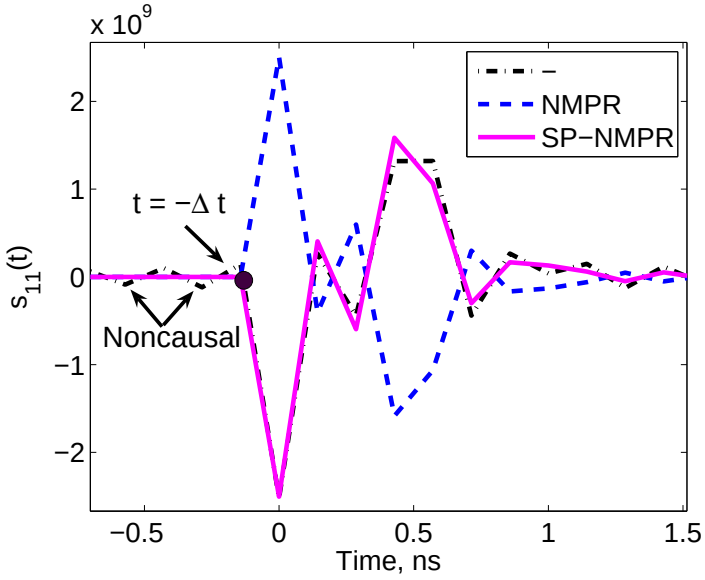



(a) $s_{11}(t)$.

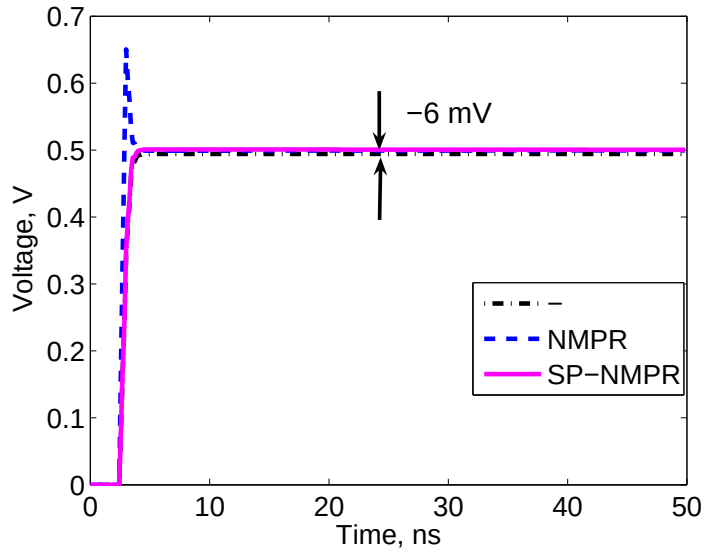



(b) Voltage at $p_1$.

**Figure 109. Comparison of impulse and step responses from different techniques with no DCLVs.**

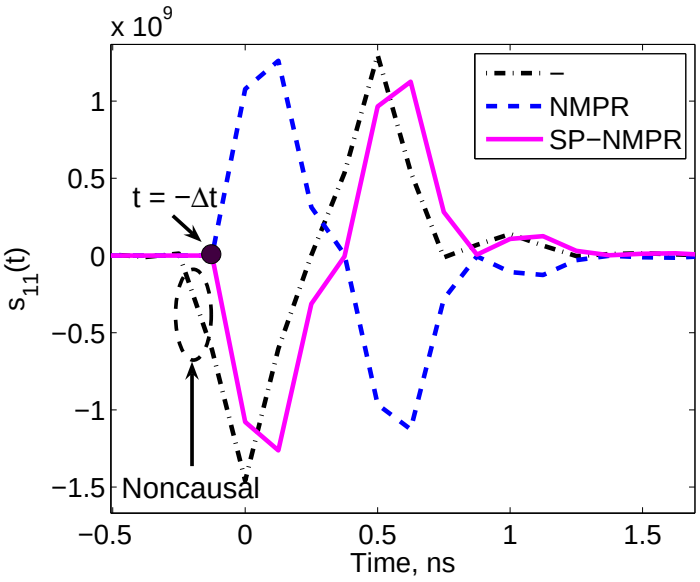



(a) $s_{11}(t)$.

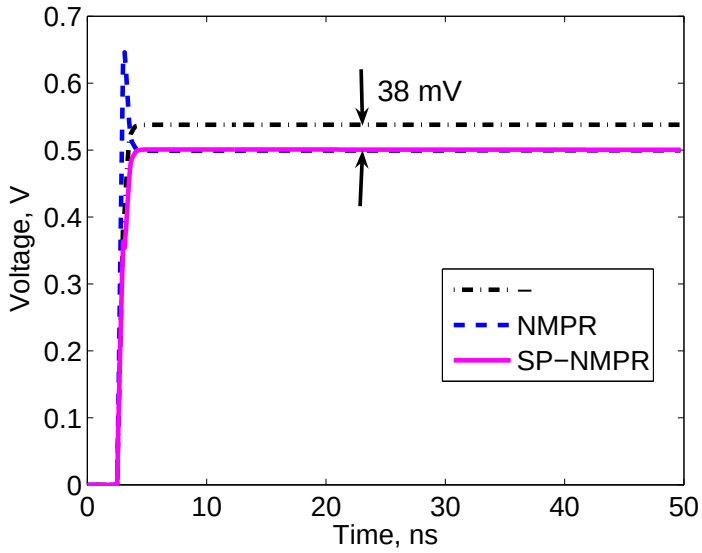



(b) Voltage at $p_1$.

**Figure 110. Comparison of impulse and step responses from different techniques with $G(f)$-induced DCLV.**




(a) $s_{11}(t)$.




(b) Voltage at $p_1$.

**Figure 111. Comparison of impulse and step responses from different techniques with $W(f)$-induced DCLV.**

are observed to be computed correctly, as such an enforcement always preserves the energy.

3. When CL enforcement is performed for $s_{11}(t)$ using the SP-NMPR technique (proposed technique), both $s_{11}(t)$ and the step responses have all the advantages of the corresponding quantities obtained without the sign preservation. However, unlike the $s_{11}(t)$ obtained without the sign preservation, correct sign of the $s_{11}(t)$ is captured only using the proposed technique (see Figure {109-111}(a)). As a result, the spurious spike in the voltage at $p_1$ (observed w/o sign preservation) is not observed using the proposed technique. The step responses from the proposed technique are also reasonable accurate even in the presence of DCLVs.

## 11.5 Summary

In this chapter, an accurate delay-causal transient simulation of interconnects characterized by b.l.f.d. data using a sign-preserving nonminimum-phase reconstruction technique has been proposed. Frequency-domain windowing has been shown to make impulse response noncausal. The proposed technique ensures causality of self impulse responses with windowing. The transient results from the proposed technique have been shown to be causal and reasonably accurate.

# CHAPTER 12
# CONCLUSIONS AND FUTURE WORK

Interconnects in modern microprocessors are not electrically ideal. This nonideality affects both performance and functionality of processors and therefore has to be studied. This study requires modeling and simulating their electrical behavior, which is not a trivial task. This dissertation is about modeling and simulation of electrical interconnects in microprocessors. This dissertation consists of two parts. The first part is about the simulation of power-supply noise (PSN) in on-chip power distribution networks (PDNs). The second part is about the cosimulation of interconnects characterized by band-limited frequency-domain data with terminations having arbitrary SPICE equivalent circuits.

Power distribution networks (PDNs) are conducting structures employed in semiconductor systems with the aim of providing circuits with reliable and constant operating voltage. This network has non-neglible electrical parasitics. Consequently, when digital circuits inside the chip switch, the supply voltage delivered to them does not remain ideal and exhibits spatial and temporal voltage fluctuations. These fluctuations in the supply voltage, known as the power-supply noise (PSN), can affect the functionality and the performance of modern microprocessors. The design of this PDN in the chip is an important part in ensuring power integrity. Modeling and simulation of the PSN in on-chip PDNs is important to reduce the cost of processors. These PDNs have irregular geometries, which affect the PSN. As a result, they have to be modeled. The problem sizes encountered in this simulation are usually large (on the order of millions), necessitating computationally efficient simulation approaches. Existing approaches for this simulation do not guarantee at least one of the following three required properties: computationally efficiency, accuracy, and numerically robustness. Therefore, there is a need to develop accurate, numerically robust, and

efficient algorithms for this simulation.

Commercial circuit simulators, based on a SPICE framework, are accurate and also numerically robust, but are not efficient for such large problem sizes. Majority of the existing methods focus on improving the efficiency in a SPICE-based framework. However, they either compromise accuracy or are not numerically robust. Minority of the existing methods are based on a finite-difference time-domain-based (FDTD-based) framework. This framework has provided an accurate, numerically robust, and computationally efficient solution for Maxwell's equations. For circuits, this framework guarantees a SPICE-like accuracy and numerical robustness. However, the computational efficiency of this framework for on-chip PSN simulation has not been studied. Moreover, these methods have only focussed on regular on-chip PDNs. This dissertation proposes using a new FDTD-based method known as the latency insertion method (LIM) to simulate PSN in on-chip PDNs. Using the proposed method, apart from accuracy and numerical robustness, even the efficiency is guaranteed in common on-chip PDN equivalent circuits. The numerical stability of LIM in irregular on-chip PDN equivalent circuits is proven.

For many passive systems (e.g., transmission lines, board connectors, package PDNs), only their frequency responses and SPICE circuits (e.g., nonlinear switching drivers, equivalent circuits of interconnects) terminating them are known. These frequency responses are usually available only up to a certain maximum frequency. Simulating the electrical behavior of these systems is important for the reliable design of microprocessors and for their faster time-to-market. Because terminations can be nonlinear, a transient simulation is required. There is a need for a transient simulation of band-limited frequency-domain data characterizing a multiport passive system with SPICE circuits. The number of ports describing the passive systems can be large ($\geq$ 100 ports). In this simulation, unlike in the traditional circuit simulators, normal properties like stability and causality of transient results are not automatically met

and have to be ensured. Existing techniques for this simulation do not guarantee at least one of the following three required properties: computationally inefficiency for a large number of ports, causality, and accuracy. Therefore, there is a need to develop accurate and efficient time-domain techniques for this simulation that also ensure causality.

Traditional circuit simulators cannot be employed for this simulation, as they deal with SPICE circuits and not with frequency responses. There are two approaches to this simulation. In the first, a SPICE equivalent circuit that has an approximate frequency response to the original frequency response is constructed, and commercial circuit simulators are employed on the combined circuit. Existing techniques based on this approach are reasonably accurate. However, this approach is not efficient when the number of ports is large. This approach is referred to as the recursive-convolution-based approach. In the second approach, frequency responses are converted to time-domain responses using IFFT, and the resulting time-domain responses are integrated with SPICE circuits, using a numerical-convolution framework. This approach, unlike the first approach, is computationally optimal with respect to the number of ports. This approach can yield reasonably accurate results, sometimes better than the results from the first approach. Existing techniques based on this approach have ensured causality, but have also introduced significant inaccuracy in the process. Also, these techniques cannot handle arbitrary SPICE circuits for terminations. In this dissertation, a new numerical-convolution-based time-domain technique is proposed that is causal and can handle arbitrary terminations. Causality enforcement procedure does not introduce the kind of inaccuracy experienced in the other existing techniques for this simulation.

## 12.1 Conclusions

Based on the work presented in Chapters 2 through 11, the contributions of this research can be listed as follows:

1. *Efficient Circuit-FDTD Method-based Formulation in Presence of Crossover Capacitance*

   The overlap capacitance between power-ground lines in adjacent metal layers of an on-chip PDN, also known as the crossover capacitance, is included in the PDN equivalent circuit. Existing circuit-FDTD based methods do not guarantee linear computational complexity per time step of the transient simulation when crossover capacitances are present. A new formulation for circuit-FDTD method has been proposed that guarantees linear computational complexity per time step of the transient simulation even in the presence of crossover capacitance. The new formulation is presented for both frequency-independent and frequency-dependent equivalent circuits of on-chip PDNs.

2. *Application of Circuit-FDTD method in Irregular PDNs*

   The circuit-FDTD method, originally only applied to regular (uniform line spacing, uniform line widths, and continuous power/ground lines running from one side of the chip to the other) on-chip PDNs, has been extended to irregular on-chip PDNs. The accuracy of the implementation has been verified through simulations. The effect of the lossy silicon substrate has been included in the simulation. As for the circuit-FDTD method, because of the irregularities in the PDN, each node in the PDN places a separate constraint on the maximum time step.

3. *Identification of Accuracy and Efficiency Issues in Performing a DC Simulation using Circuit-FDTD Method*

A new problem has been identified when the circuit-FDTD method is applied to DC simulation. This problem concerns the accuracy of the DC node voltages computed. It has been found that when DC simulation is computed using the circuit-FDTD method, the oscillations from step responses do not settle down (or die down) in some nodes. When transient PSN simulation is started on a circuit with unsettled step responses, the PSN can be computed inaccurately in two ways: 1) The PSN can have contributions not only from switching currents (which it should) but also from unsettled step responses (which it should not). 2) The PSN can be observed in a location even before the effect of the switching current can be felt at the location, i.e., the PSN computation can violate causality. This new problem has been solved by running the DC simulation for sufficiently long time so that the step responses are significantly settled in all nodes. Unfortunately, in the modified simulation, it has been observed that the DC simulation took majority of the total simulation time.

4. *On-Chip Power Grid Simulation using LIM*

   Circuit-FDTD method guarantees linear computational complexity per time step of the transient simulation only in circuits where there is latency in every node and branch of the circuit. It is shown, however, that this latency requirement may not be met in equivalent circuits of on-chip PDNs. To preserve the computational complexity, it has been proposed to insert artificial latency in missing places of the circuit. The circuit-FDTD method augmented with artificial latency is referred to as the latency insertion method (LIM). LIM, like any FDTD-based method, is only conditionally stable. The time step of the transient simulation cannot be arbitrary and depends on the smallest inductance-capacitance in the circuit. Care has to be taken about the values of artificial latency elements. If the artificial element values are too large, then the accuracy can be significantly affected. On the other hand, if these element

values are too small, then time step of the transient simulation has to be made small. Unlike LIM, which rely on several (transient simulation) iterations to compute the latency elements, this dissertation proposes closed-form expressions for computing the latency elements. These expressions take into account the element values, the maximum frequency in the excitation, and the accuracy required. Therefore, time step can be made just small enough to meet the accuracy requirements. Unlike in LIM, simulation need not be repeated for accuracy. The LIM-enabled power grid transient simulator is demonstrated to be as accurate and robust as SPICE, to have linear time complexity per time step of the transient simulation, and to have linear memory complexity for the whole transient simulation. The total number of time steps required in this simulator is shown to be $O(N_n^{1-1.5})$ for practical on-chip power grid problems.

5. *On-Chip LIM Including On-Chip Decoupling Capacitors and Package Parasitics*

   LIM has been extended to simulate power-supply noise in on-chip power grids in the presence of on-chip decoupling capacitors. An RC model has been used for the on-chip decoupling capacitance. To retain optimal memory and time complexity per time step of the simulation, to each on-chip decap, a fictitious inductance has been inserted. The accuracy of the simulation has been verified against SPICE. The effect of on-chip decoupling capacitance on the power-supply noise has been demonstrated.

   LIM has also been extended to simulate power-supply noise in on-chip power grids in the presence of package parasitics. The package has thus far been modeled as an ideal voltage source. As a first-level model, the C4 bump and the package has been modeled as a series RL branch. This branch is put in C4 locations. The value of the resistance and inductance can be obtained from the input impedance seen from C4 terminals to the end of the package. The

computational complexity of the LIM has been retained. The accuracy of this simulation is verified against SPICE. The importance of modeling package PDNs even while simulating PSN in power grids is verified through simulations.

6. *Effect of the On-Chip Inductance on PSN*

   Using the proposed formulation, the effect of on-chip inductance on the power-supply noise has been studied. It has been found that the on-chip inductance has three effects that potentially affect the PSN computation: 1) On-chip inductance lowers the frequency of the chip-package resonance. 2) On-chip inductance lowers the magnitude of the peak impedance, usually observed near the chip-package resonant frequency. 3) On-chip inductance introduces new resonances at frequencies greater than the chip-package resonant frequency. These extra resonances introduce a fast variation to the power supply in the time domain. This variation can make the power supply fluctuate beyond the restricted margin, although this violation is temporary. This sudden variation in power supply would not be captured if the on-chip inductance is not modelled as part of the power grid simulation.

7. *Analytical Stability Conditions of LIM in Inhomogeneous RLC and GLC Circuits*

   LIM, unlike the SPICE-based approaches, is not guaranteed to be stable when there are discontinuities in the circuits. Until now, it has not been possible to prove the stability of LIM for inhomogeneous circuits. The stability of LIM has been proven for inhomogeneous RLC and GLC circuits. With this proof, the proof for stability of LIM-enabled power grid simulation for irregular power grids is established for cases where capacitive coupling can be ignored.

8. *Conditional Stability of Alternate Direction implicit Methods*

Alternate direction implicit (ADI) method has been used to relax the time step of the transient simulation using the transmission line method (TLM), an explicit method similar to the circuit-FDTD method. It was found that 1) the ADI method can only be applied to mesh-type equivalent circuits (where two orthogonal directions of propagation are possible in every metal layer) and 2) the ADI method for mesh-type equivalent circuits becomes unstable for some choices of time step when open-circuit boundary conditions are applied at the circuit boundary. Therefore, it has been concluded that the ADI method cannot be used to relax the time step of the circuit-FDTD for the on-chip PDN equivalent circuits considered in this research.

9. *Causality Enforcement Using Minimum-Phase/All-Pass Decomposition*

   When band-limited multiport frequency-domain data are present, the causality of the multiport impulse responses are ensured traditionally as follows: the multiport impulse responses are computed numerically using IFFT, and the causality of the responses is enforced in the time domain by truncating each port-to-port impulse response before the corresponding port-to-port propagation delay. It has been shown that such truncation-based causality enforcement techniques do not preserve the energy of the individual frequency responses and can result in inaccurate transient results. To avoid this drawback, a new causality enforcement technique has been proposed. In this technique, the multiport frequency responses are causally reconstructed in the frequency domain using a minimum-phase/all-pass decomposition of the responses, and the reconstructed frequency responses are converted to the corresponding impulse responses numerically using the IFFT. It has been observed that the new technique does not suffer from the inaccuracy issues observed in the truncation-based techniques.

10. *Causal Transient Simulation of Band-Limited Frequency-Data with SPICE Circuits*

    A new causal transient simulation engine that integrates band-limited frequency domain data characterizing a multiport linear system with SPICE circuits has been proposed. This integration has been achieved by formulating a numerical convolution-based approach in a MNA framework. The advantage of this engine are that the port terminations can be arbitrary and the transient results are causal. The accuracy of the transient simulation has been verified for frequency-domain data characterizing transmission lines.

11. *Sign-Preserving Minimum-Phase/All-Pass Decomposition*

    Using the causality enforcement technique discussed thus far, the leading signs of the frequency responses are not preserved consistently. Not preserving this sign can make the transient results inaccurate. This nonpreservtion of sign is because of the existing functional form of the all-pass component. Using this form, the leading negative sign of the frequency response cannot be modeled. To capture the leading sign, a constant sign term has been included as part of the all-pass component. The accuracy of the new functional form of the all-pass component, and the accuracy of the causal transient results using this decomposition have been demonstrated.

12. *Causality Enforcement for Self Frequency Responses*

    Thus far, causality of the transient results has been ensured by causally reconstructing the transfer frequency responses, i.e., frequency response between two different ports. However, the self responses are not reconstructed and are not converted to time domain using IFFT. It has been shown that conventional way of treating self responses implicitly truncates the self impulse responses for $t < 0$. This truncation affects the accuracy. To overcome this

inaccuracy, even the self responses are reconstructed using the sign-preserving minimum-phase/all-pass decomposition. Subsequent IFFT of the reconstructed self responses yields a causal self impulse response without the inaccuracy issues related to the truncation-based technique. The accuracy of the reconstruction and of the transient results are demonstrated.

13. *Frequency-Domain Windowing Induces Causality Violations*

    In all numerical-convolution-based approaches, the band-limited frequency-domain data are usually subjected to a frequency-domain windowing for making the transient results smooth and stable (sometimes). The strength of the window is chosen based on accuracy and stability considerations. In this dissertation, it has been shown, for the first time, that frequency-domain windowing makes causal frequency-domain data noncausal. It has been also demonstrated that the bigger the strength of windowing, the larger the noncausality the data become. As a result, when frequency-domain windowing is applied, the transient results are not going to be causal unless ensured.

    Band-limited nature of data can be considered as applying a rectangular window to the data. However, this windowing does not make the data causal, instead it only makes the time response noncausal. This noncausality is not so serious as the noncausality from other windowing.

## 12.2 Future work

The possible future work in the on-chip power grid simulation is the following.

1. *Efficient Methods to Improve the Time Complexity of LIM-enabled Power Grid Simulation*

    One of the drawbacks of the LIM-enabled power grid simulation is the small time step needed for the transient simulation. Because of the small time step,

the number of time steps required in the transient simulation runs into millions for most practical problems. Usually, the number of time steps should not be that large, especially given that the size of the problem is also of the same order. Therefore, there is a clear need to improve the time complexity of the LIM-enabled power grid simulation.

Since the source of the problem is the small time step, methods to relax the (strict) constraint on the time step are necessary. Similar problem has already been addressed in FDTD methods for solving Maxwell's equations [76], [77]. The approach used is based on alternate direction implicit (ADI) methods. Some effort has already been done in applying an ADI-based scheme for the power grid problem in [43], [38]. However, these efforts have not been successful. A good starting point is to address the shortcomings of the ADI implementation described in [43], [38].

2. *Accurate Equivalent Circuits and Parasitic Extraction*

   As was already mentioned in Chapter 1, a overwhelming majority of the prior work in the power grid simulation focus only making the DC or transient simulation efficient. An integral part of the simulation is the accuracy of the equivalent circuit and the parasitic extraction. However, the latter issue has not received much attention. Some effort in this direction has already been taken in [31]. A good starting point is first to integrate the extraction engine proposed in [31] with the simulation engine proposed in this dissertation and then study the accuracy of the extraction and the equivalent circuit.

3. *Analytical Stability Conditions of LIM for Inhomogeneous RLC Circuits Containing Coupling Capacitance*

   Establishing analytical stability conditions of LIM for a circuit to be simulated is necessary to evaluate the method's overall efficiency. On-chip power grid

equivalents proposed in this dissertation have branch (coupling) capacitances in the form of on-chip decoupling capacitance and crossover capacitance. However, analytical stability conditions of LIM have been established for these equivalent circuits only in the absence of branch capacitances. Therefore, establishing analytical stability condition of LIM in the presence of branch capacitors is still a open problem and is necessary too.

4. *Chip-Package Cosimulation*

   The need to include the effect of the package when simulating the chip PDN has become stronger nowadays. Many design companies have set up special teams that focus on chip-package codesign in their product development. The CAD community have only addressed this cosimulation problem at a post-layout level. Not much attention has been paid for this problem at the pre-layout level. Some effort in this direction include [45], [7], [79], [80] and this dissertation.

The possible future work in the transient simulation of band-limited data are the following.

1. *Study of the Effect of Minimum-Phase Reconstruction-based Causality Enforcement Schemes on the Accuracy of Transient Simulation*

   One of the drawbacks of the minimum-phase reconstruction-based causality enforcement scheme proposed in this dissertation is that it reconstructs the phase of a frequency response from the amplitude of the response. The phase of the frequency response affects the accuracy of the simulation. However, not much effort has been done on studying the accuracy of the phase reconstruction and the effect of the phase inaccuracy on the accuracy of the transient simulation.

2. *Study of the Effect of the Nature of Band-limited Data on the Accuracy of Numerical-Convolution-Based Transient Simulation*

The accuracy of numerical-convolution-based transient simulation of band-limited data has not been studied properly. The accuracy of this simulation depends on the data. This dissertation has only focussed on the data from transmission lines. To propose numerical-convolution-based transient simulation for non-transmission line problems (e.g., board connectors, wire bonds, C4 bumps), the accuracy of this simulation for generic data has to be studied. Not much work has been done towards this end.

## 12.3 Publications

The following publications have resulted during the course of this research:

1. **S. N. Lalgudi**, J. Mao, and M. Swaminathan, "Parasitic Extraction and Simulation of Simultaneous Switching Noise in On-Chip Power Distribution Networks," *IEEE Conference on Electromagnetic Compatibility,* Mar. 2005, Zurich.

2. **S. N. Lalgudi**, M. Swaminathan and Y. Kretchmer, "Simulation of Simultaneous Switching Noise in On-Chip Power Distribution Networks of FPGAs," *IEEE 14th Topical Meeting on Electrical Performance of Electronic Packaging*, Oct. 2005, pp. 319-322.

3. **S. N. Lalgudi**, K. Srinivasan, G. Casinovi, R. Mandrekar, E. Engin, M. Swaminathan, and Y. Kretchmer, "Causal Transient Simulation of Systems Characterized by Frequency-Domain Data in a Modified Nodal Analysis Framework," *IEEE 15th Topical Meeting on Electrical Performance of Electronic Packaging*, Oct. 2006.

4. S. Mukherjee, M. Swaminathan, and **S. N. Lalgudi**, "Broadband Modeling and Tuning of Multi-layer RF Circuits using Physical Augmentation Methodology," *IEEE Asia Pacific Microwave Conference*, Sept. 2007.

5. **S. N. Lalgudi**, M. Swaminathan, and Y. Kretchmer, "On-Chip Power Grid Simulation using Latency Insertion Method," Accepted for Publication in *IEEE Trans. on Circuits and Systems-I: Fundamental theory and applications*, Vol. 55, No. 3, April 2008.

6. **S. N. Lalgudi**, E. Engin, G. Casinovi, and M. Swaminathan , "Accurate Transient Simulation of Interconnects Characterized by Band-Limited Data With Propagation Delay Enforcement in a Modified Nodal Analysis Framework," Accepted for Publication in *IEEE Trans. on Electromagnetic Compatibility*, July 2008.

7. **S. N. Lalgudi**, and M. Swaminathan, "Analytical Stability Condition of the Latency Insertion Method for Inhomogeneous GLC Circuits," Accepted for Publication in *IEEE Trans. on Circuits and Systems-II: Express Briefs*, 2008.

# APPENDIX A

# FDTD METHOD FOR SOLVING MAXWELL'S EQUATIONS

*The FDTD method for Solving Full-Wave Problems*

The FDTD method is a time-domain technique to solve Maxwell's equations. The objective in this method is to compute spatial and temporal profiles of electric and magnetic fields. The Maxwell's curl equations involving electric fields and involving magnetic fields are solved. This solution is performed in an unique manner that guarantees a high accuracy and an optimal computational complexity. These properties are ensured by a careful choice of the kind and type of integration rule. Standard central differencing is usually employed, and an Yee-like gridding is incorporated. The nonzero propagation delay of the wave between any two points in the FDTD grid facilitates applying Yee gridding. As a result of this integration rule, the accuracy of the method scales as $O\left((\Delta x)^2\right)$ in space and as $O\left((\Delta t)^2\right)$ in time. The formulation ensures that a diagonal matrix is solved for most of the applications. As a result, the memory complexity of the simulation and the time complexity per time step of the simulation are $O(N_n)$ each. This technique has been successfully applied even to media with nonuniform ($\equiv$ irregular) material properties. The FDTD method, however, suffers from the following problems:

1. The biggest problem is that the transient simulation is only conditionally stable, i.e., for a stable result, the time step of the transient simulation *cannot* be any arbitrary positive real number less than the total simulation time. The upper bound for the time step is predicted to be less than the time taken for the wave to travel one unit cell in the discretized problem domain. The size of the unit cell may be too small, necessitated by the need to model a fine feature in the medium. This upper bound can be lot smaller than the smallest rise time in

the excitation. As a result of this small time step, the number of time steps, $N_t$, can be large. Large $N_t$ affects the time complexity of the overall transient simulation.

2. The formulation of the FDTD method depends on the constitutive relations of the medium. This means the update expressions for electric and magnetic fields depend on type of the medium. The computational complexity of the formulation has been ensured to be optimal for different types of media properties: homogeneous or inhomogeneous, isotropic or anisotropic, lossless or lossy, dispersive or nondispersive, etc.

3. Though the method has been experimentally observed to be working for materials with nonuniform material properties, not all of these observations have been proven.

# REFERENCES


[1] D. J. Herell and B. Beker, "Modeling of power distribution systems for high-performance microprocessors," *IEEE Trans. on Advanced Packaging*, vol. 22, pp. 240–248, August 1999.

[2] M. Swaminathan, J. Kim, I. Novak, and J. P. Libous, "Power distribution networks for system-on-package: Status and challenges," *IEEE Trans. on Advanced Packaging*, vol. 27, pp. 286–300, May 2004.

[3] L. W. Nagel, "Spice2, a computer program to simulate semiconductor circuits," tech. rep., University of California, Berkeley, CA, 1975.

[4] G. A. Katopis, "Delta-i noise specification for a high-performance computing machine," *Proceedings of the IEEE*, vol. 73, pp. 1405–1415, September 1985.

[5] H. B. Bakoglu, *Circuits, interconnections, and packaging for VLSI.* MA: Addison-Wesley Publishing Company, 1990.

[6] K. L. Shepard and V. Narayanan, "Noise in deep submicron digital design," in *IEEE/ACM International Conference on Computer-Aided Design*, pp. 10–14, November 1996.

[7] H. H. Chen and J. S. Neely, "Interconnect and circuit modeling techniques for full-chip power supply noise analysis," *IEEE Trans. on Components, Packaging, and Manufacturing Technology - Part B*, vol. 21, pp. 209–215, August 1998.

[8] L. H. Chen, M. Marek-Sadowska, and F. Brewer, "Coping with buffer delay change due to power and ground noise," in *Proceedings on Design Automation Conference*, pp. 860–865, 2002.

[9] R. Ahmadi and F. N. Najm, "Timing analysis in presence of power supply and ground voltage variations," in *IEEE/ACM Internation Conference on Computer-Aided Design of Integrated Circuits*, pp. 176–183, November 2003.

[10] R. Saleh, S. Z. Hussain, S. Rochel, and D. Overhauser, "Clock skew verification in the presence of ir-drop in the power distribution network," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, pp. 635–644, June 2000.

[11] M. Saint-Laurent and M. Swaminathan, "Impact of power-supply noise on timing in high-frequency microprocessors," *IEEE Trans. on Advanced Packaging*, vol. 27, pp. 135–144, Febraury 2004.

[12] W. S. Song and L. A. Glasser, "Power distribution techniques for vlsi circuits," *IEEE Journal on Solid-State Circuits*, vol. 21, pp. 150–156, February 1986.

[13] P. E. Gronowski, W. J. Bowhill, R. P. Preston, M. K. Gowan, and R. L. Allmon, "High-performance microprocessor design," *IEEE Journal on Solid-state Circuits*, vol. 33, pp. 676–686, May 1998.

[14] P. Larsson, "Resonance and damping in cmos circuts with on-chip decoupling capacitors," *IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 45, pp. 849–858, August 1998.

[15] S. Bobba, T. Thorp, K. Aingaran, and D. Liu, "Ic power distribution challenges," in *IEEE/ACM Internation Conference on Computer-Aided Design of Integrated Circuits*, pp. 643–650, November 2001.

[16] K. Aygun, M. J. Hill, K. Eilert, K. Radhakrishnan, and A. Levin, "Power delivery for high-performance microprocessors," *Intel Technology Journal*, vol. 9, pp. 273–284, November 2005.

[17] "Power grid verification," tech. rep., Cadence Design Systems, Inc., San Jose, CA, 2002.

[18] K. Roy, S. Mukhopadhaya, and H. Mahmoodi-Meinand, "Leakage current mechanism and leakage reduction techniques in deep-submicronmeter cmos circuits," *Proceedings of IEEE*, vol. 9, pp. 305–327, February 2003.

[19] N. S. Kim, T. Austin, D. Blaauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage current: Moore's law meets static power," *Computer*, vol. 36, pp. 68–75, December 2003.

[20] T. Kam, S. Rawat, D. Kirkpatrick, R. R. Roy, G. S. Spirakis, N. Sherwani, and C. Peterson, "Eda challenges facing future microprocessor design," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, pp. 1498–1506, December 2000.

[21] I. A. Ferzli and F. N. Najm, "Analysis and verification of power grids considering process-induced leakage-current variations," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, pp. 126–143, January 2006.

[22] R. Downing, P. Gebler, and G. Katopis, "Ieee trans. on components, hybrids, and manufacturing technology," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 16, pp. 484–489, August 1993.

[23] J. Prymak, "Advanced decoupling using ceramic mlc capacitors," in *IEEE 40th Proc. of Electronic Components and Technology Conference*, pp. 1014–1023, May 1990.

[24] W. D. Becker, K. Eckhardt, R. W. Frech, G. A. Katopis, E. Klink, M. F. McAllister, T. G. McNamara, P. Muench, S. R. Richter, and H. H. Smith, "Modeling, simulation and measurement of mid-frequency simultaneous switching noise in computer systems," *IEEE Trans. on Advanced Packaging*, vol. 21, pp. 157–163, May 1998.

[25] L. D. Smith, "Decoupling capacitor calculations for cmos circuits," in *IEEE 3rd Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 101–105, November 1994.

[26] L. D. Smith, R. E. Anderson, D. W. Forehand, T. J. Pelc, and T. Roy, "Power distribution system design methodology and capacitor selection for modern cmos technology," *IEEE Trans. on Advanced Packaging*, vol. 22, pp. 284–291, August 1999.

[27] A. Dharchoudhury, R. Panda, D. Blaauw, and R. Vaidyanathan, "Design and analysis of power distribution networks in powerpc$^{TM}$ microprocessors," in *Proceedings on Design Automation Conference*, pp. 738–743, June 1998.

[28] L. D. Smith, R. Anderson, and T. Roy, "Power plane spice models and simulated performance for materials and geometries," *IEEE Trans. on Advanced Packaging*, vol. 24, pp. 277–287, August 2001.

[29] J.-H. Kim and M. Swaminathan, "Modeling of multilayered power distribution planes using transmission matrix method," *IEEE Trans. on Advanced Packaging*, vol. 25, pp. 189–199, May 2002.

[30] J. N. Kozhaya, S. R. Nassif, and F. N. Najm, "A mutigrid-like technique for power grid analysis," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, pp. 1148–1160, October 2002.

[31] J. Mao, *Modeling of simultaneous switching noise in multilayered electronic packaging and integrated circuits.* PhD thesis, Dept. of Elect. and Comput. Engg., Georgia Institute of Technology, Atlanta, USA, 2004.

[32] M. Zhao, R. V. Panda, S. S. Sapatnekar, and D. Blaauw, "Hierarchical analysis of power distribution networks," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 21, pp. 159–168, February 2002.

[33] E. Chiprout, "Fast flip-chip power grid analysis via locality and grid shells," in *IEEE/ACM Conference on Computer-Aided Design of Integrated Circuits and Systems*, pp. 485–488, November 2004.

[34] H. Qian, S. R. Nassif, and S. S. Sapatnekar, "Power grid analysis using random walks," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, pp. 1204–1224, August 2005.

[35] T. H. Chen and C. C. P. Chen, "Efficient large-scale power grid analysis based on preconditioned krylov-subspace iterative methods," in *Proceedings on Design Automation Conference*, pp. 559–562, August 2001.

[36] Y. Zhong and M. D. F. Wong, "Fast algorithms for ir drop analysis in large power grid," in *IEEE conference on Computer-Aided Design of Integrated Circuits and Systems*, pp. 351–357, November 2005.

[37] W. Guo, S. X. Tan, Z. Luo, and X. Hong, "Partial random walk for large linear network analysis," in *Proceedings of the International Symposium on Circuits and Systems*, pp. 173–176, November 2004.

[38] W. Guo and S. X. D. Tan, "Circuit-level alternating-direction implicit approach to transient analysis of power distribution networks," in *International Conference on Application-Specific Integrated Circuits*, vol. 1, pp. 246–249, October.

[39] J. Choi, L. Wan, M. Swaminathan, B. Beker, and R. Master, "Modeling of realistic on-chip power grid using the fdtd method," in *IEEE International Symposium on Electromagnetic Compatibility*, pp. 238–243, August 2002.

[40] J. Choi, M. Swaminathan, N. Do, and R. Master, "Modeling of power supply noise in large chips using the circuit-based finite-difference time-domain method," *IEEE Trans. on Electromagnetic Compatibility*, vol. 47, pp. 424–439, August 2005.

[41] J. Mao, M. Swaminathan, J. Libous, and D. O'Connor, "Effect of substrate resistivity on switching noise in on-chip power distribution networks," in *IEEE 12th Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 33–36, November 2003.

[42] S. N. Lalgudi, Y. Kretchmer, and M. Swaminathan, "Simulation of simultaneous switching noise in on-chip power distribution networks of fpgas," in *IEEE 14th Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 319–322, October 2005.

[43] Y. M. Lee and C. C. P. Chen, "The power grid transient simualation in linear time based on 3-d alternating-direction-implicit method," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 22, pp. 1545–1550, November 2003.

[44] L. T. Pillage, R. A. Rohrer, and C. Visweswariah, *Electronic circuit and system simulation methods*. NY: McGraw-Hill Inc, 2 ed., 1994.

[45] J. Choi, *Modeling of power supply noise in large chips using the finite difference time domain method*. PhD thesis, Dept. of Elect. and Comput. Engg., Georgia Institute of Technology, Atlanta, USA, 2002.

[46] K. S. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Trans. on Antennas and Propagation*, vol. 14, pp. 302–307, May 1996.

[47] C. R. Paul, "Incorporation of terminal constraints in the fdtd analysis," *IEEE Trans. on Electromagnetic Compatibility*, vol. 36, pp. 85–91, May 1994.

[48] J. E. Schutt-Aine, "Latency insertion method (lim) for the fast transient simulation of large networks," *IEEE Trans. on Circuits and Systems-I: Fundamental theory and applications*, vol. 48, pp. 81–88, January 2001.

[49] Z. Deng and J. Schutte-Aine, "Stability analysis of latency insertion method (lim)," in *IEEE Proc. of Electronic Components and Technology Conference*, pp. 1866–1873, June 2004.

[50] A. George, "Nested dissection of a regular finite difference mesh," *SIAM Journal of Numerical Analysis*, vol. 10, pp. 345–363, April 1973.

[51] H. M. Nussenzveig, *Causality and dispersion relations.* Academic Press, Inc., 1972.

[52] P. Triverio, S. Grivet-Talocia, M. S. Nakhla, F. G. Canavero, and R. Achar, "Stablity, causality, and passivity in electrical interconnect models," *IEEE Trans. on Advanced Packaging*, vol. xx, pp. 1–14, xx 2007.

[53] S. Grivet-Talocia and A. Uboli, "On the generation of large passive macromodels for complex interconnect structures," *IEEE Trans. on Advanced Packaging*, vol. 29, pp. 39–54, February 2006.

[54] P. Triverio and S. Grivet-Talocia, "On checking causality of bandlimited sampled frequency responses," in *15th IEEE Conf. on Elect. Performance of Electron. Packag.*, vol. November, pp. 501–504.

[55] R. Mandrekar and M. Swaminathan, "Causality enforcement in transient simulation of passive networks through delay extraction," in *9th IEEE Workshop on Signal Propag. Interconnects*, pp. 25–28, May.

[56] R. Mandrekar, K. Srinivasan, E. Engin, and M. Swaminathan, "Causal transient simulation of passive networks with fast convolution," in *10th IEEE Workshop on Signal Propag. Interconnects*, pp. 61–64, May.

[57] R. Mandrekar, *Modeling and cosimulation of signal distribution and power delivery in packaged digital systems.* PhD thesis, Dept. of Elect. and Comput. Engg., Georgia Institute of Technology, Atlanta, USA, 2006.

[58] S. Lin and E. S. Kuh, "Transient simulation of lossy interconnects based on a recursive convolution formulation," *IEEE Trans. on Circuits Syst. I: Fundam. Theory Appl.*, vol. 39, pp. 879–892, November 1992.

[59] S. Grivet-Talocia, H. M. Huang, A. E. Ruehli, F. Canavero, and I. M. Elfadel, "Transient analysis of lossy transmission lines: An efficient approach based on the method of characteristics," *IEEE Trans. on Adv. Packag.*, vol. 27, pp. 45–56, February 2004.

[60] S. Min, *Automated construction of macromodels from frequency data for simulation of distributed interconnect networks*. PhD thesis, Dept. of Elect. and Comput. Engg., Georgia Institute of Technology, Atlanta, USA, 2004.

[61] A. Dounavis, N. Nakhla, R. Achar, and M. Nakhla, "Delay extraction and passive macromodeling of lossy coupled transmission lines," in *IEEE Conf. on Elect. Performance of Electron. Packag.*, vol. 1, pp. 251–254, October.

[62] A. R. Djordevic and T. K. Sarkar, "Transient analysis of electromagnetic systems with multiple lumped nonlinear loads," *IEEE Trans. on Antennas Propag.*, vol. 33, pp. 533–539, May 1985.

[63] D. Winklestein, M. B. Steer, and R. Pomerleau, "Simulation of arbitrary transmission line networks with nonlinear terminations," *IEEE Trans. on Circuits Syst.*, vol. 38, pp. 418–422, April 1991.

[64] L. P. Vakanas, A. C. Cangellaris, and O. A. Palusinski, "Scattering parameter-based simulation of transients in lossy nonlinearly terminated packaging interconnections," *IEEE Trans. on Compon. Packag. Manuf. Technol. B*, vol. 17, pp. 472–479, November 1994.

[65] M. S. Basel, M. B. Steer, and P. D. Franzon, "'simulation of high speed interconnects using a convolution-based hierarchical simulator," *IEEE Trans. on Compon. Packag. Manuf. Technol. B*, vol. 18, pp. 74–82, February 1995.

[66] J. Mao, M. Swaminathan, J. Libous, , and D. O'Connor, "Effect of substrate resistivity on switching noise in on-chip power distribution networks," pp. 33–36, 2003.

[67] A. Scarlatti and C. L. Holloway, "An equivalent transmission-line model containing dispersion for high-speed digital lines - with an fdtd implementation," *IEEE Trans. on Electromagnetic Compatibility*, vol. 43, pp. 504–514, November 2001.

[68] B. Gustavsen and A. Semlyen, "Rational approximation of frequency domain responses by vector fitting," *IEEE Transactions on Power Delivery*, vol. 14, pp. 1052–1061, July 1999.

[69] A. Taflove and S. Hagness, *Computational Electrodynamics*. MA: Artech House Inc., 2000.

[70] C. R. Paul, *Multiconductor transmission lines*. New Delhi, India: John Wiley & Sons, 1994.

[71] B. C. Kuo, *Automatic Control Systems, A design perspective.* Prentice-Hall, 1989.

[72] Y. Saad, *Iterative methods for linear systems.* 2 ed., 2000.

[73] Z. Panda, S. Sundareswaran, and D. Blaauw, "Impact of low-impedance substrate on power supply integrity," in *IEEE Conference on Design and Test of Computers*, vol. 20, pp. 16–22, May-June.

[74] Z. He, M. Celik, and L. Pileggi, "Spie: Sparse partial inductance extraction," in *Proceedings on Design Automation Conference*, pp. 137–140, June.

[75] A. Devgan, H. Ji, , and W. Dai, "How to efficiently capture on-chip inductance effects: Introducing a new circuit element k," in *IEEE conference on Computer-Aided Design of Integrated Circuits and Systems*, pp. 150–155, November.

[76] T. Namiki and K. Ito, "A new fdtd algorithm free from the cfl condition restraint for a 2d-te wave," in *Dig. 1999 Antennas and Propagation Symposium*, pp. 192–195, July.

[77] T. Namiki, "A 3-d adi-fdtd method - unconditionally stable time-domain algorithm for solving vector maxwell's equations," *IEEE Trans. Microwave Theory and Techniques*, vol. 48, pp. 1743–1748, October 2000.

[78] C. Ashcroft and J. W. H. Liu, "Robust ordering of sparse matrices using multisection," *SIAM Journal of Matrix Analysis and Applications*, vol. 19, no. 3, pp. 816–832, 1998.

[79] S. Pant and E. Chiprout, "Power grid physics and implications for cad," vol. 13, (CA, USA), pp. 199–204, July 2006.

[80] S. Pant, D. Blaauw, and E. Chiprout, "Power grid physics and implications for cad," *IEEE Trans. on Design and Test of Computers*, vol. 24, pp. 246–254, May-June 2007.

[81] W. L. Brogan, *Modern Control Theory.* NJ: Prentice Hall, 1992.

[82] W. Thiel and L. P. B. Katehi, "Some aspects of stability and numerical dissipation of the finite-difference time-domain (fdtd) technique including passive and active lumped elements," *IEEE Trans. Microwave Theory and Techniques*, vol. 50, pp. 2159–2165, September 2002.

[83] F. Edelvik, R. Schuhmann, and T. Weiland, "A general stability analysis of fit/fdtd applied to lossy dielectrics and lumped elements," *International Journal of Numerical modelling:Electronic Networks, Devices, and Fields*, vol. 17, pp. 407–419, 2004.

[84] F. Kung and H. T. Chuah, "Stability of classical finite-difference time-domain (fdtd) formulation with nonlinear elements - a new perspective," in *Progess in Electromagnetics Research, PIER*, vol. 42, pp. 49–89.

[85] G. Strang, *Linear Algebra and its Applications*. NJ: homson Learing Inc., 3 ed., 1998.

[86] S. N. Lalgudi, K. Srinivasan, G. Casinovi, R. Mandrekar, E. Engin, Y. Kretchmer, and M. Swaminathan, "Causal transient simulation of systems characterized by frequency-domain data in a modified nodal analysis framework," in *15th IEEE Conf. on Elect. Performance of Electron. Packag.*, pp. 123–126, November.

[87] F. M. Tesche, "On the use of the hilbert transform for processing measured cw data," *IEEE Trans. Electromagn. Compat.*, vol. 34, pp. 259–266, August 1992.

[88] W. T. Beyene and C. Yuan, "An accurate transient analysis of high-speed package interconnects using convolution technique," in *Analog Integrated Circuits and Signal Processing*, no. 35, pp. 107–120, 2003.

[89] A. Oppenheim and R. Schafer, *Discrete-time signal processing*. Prentice Hall., 2 ed., 1999.

[90] H. Curtins and A. V. Shah, "Pulse behavior of transmission lines with dielectric losses," *IEEE Trans. on Circuits Syst.*, vol. 32, pp. 819–826, August 1985.

[91] 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*, vol. 21, pp. 105–114, February 1998.

[92] D. Saraswat, R. Achar, and M. Nakhla, "Global passivity enforcement algorithm for macromodels of interconnect subnetworks characterized by sampled data," *IEEE Trans. on Very Large Scale Integr. (VLSI) Syst.*, vol. 13, pp. 819–832, July 2005.

[93] S. L. Hahn, *Hilbert transforms in signal processing*. Artech House Publications, 2 ed., 1996.

[94] T. K. Sarkar, "Generation of nonminimum phase from amplitude-only data," *IEEE Trans. on Microw. Theory Tech.*, vol. 46, pp. 1079–1084, August 1998.

[95] D. M. Pozar, *Microwave Engineering*. John Wiley and Sons, 3 ed., 2005.

[96] J. E. Schutt-Aine and R. Mittra, "Scattering parameter transient analysis of transmission lines loaded with nonlinear terminations," *IEEE Trans. on Microw. Theory Tech.*, vol. 36, pp. 529–536, November 1988.

[97] H. Chung-Wen, A. E. Ruehli, and P. A. Brennan, "The modified nodal analysis approach to network analysis," *IEEE Trans. on Circuits and Systems*, vol. 22, pp. 504–509, June 1975.

[98] K. S. Kundert, *The Designer's Guide to SPICE & SPECTRE*. Kluwer Academic Publishers, 1995.

[99] Agilent Technologies, Santa Rosa, CA, *Agilent Advanced Design System User's Guide*, 2006.

[100] Synopsys, USA, *HSPICE Simulation and Analysis User Guide*, version y-2006.03 ed., March 2006.

[101] J. R. James and G. Andrasic, "Causal transient simulation of systems characterized by frequency-domain data in a modified nodal analysis framework," in *IEEE Proceedings. on Microw., Antennas and Propagtn.*, vol. 137, pp. 184–188, June 1990.

[102] T. R. Arabi, A. T. Murphy, and T. K. Sarkar, "An efficient technique for the time-domain analysis of multi-conductor transmission lines using the hilbert transform," in *IEEE MTT-S Digest*, vol. 1, pp. 185–188, June 1992.

[103] T. R. Arabi and R. Suarez-Gartner, "Time domain analysis of lossy multiconductor transmission lines using the hilbert transform," in *IEEE MTT-S Digest*, vol. 2, pp. 987–990, October 1993.

[104] S. M. Narayana, G. Rao, R. Adve, T. K. Sarkar, V. C. Vannicola, M. C. Wicks, and S. A. Scott, "Interpolation/extrapolation of using the hilbert transform," *IEEE Trans. Electromagn. Compat.*, vol. 14, pp. 1621–1627, October 1996.

[105] P. Triverio and S. Grivet-Talocia, "Causality-constrained interpolation of tabulated frequency responses," in *15th IEEE Conf. on Elect. Performance of Electron. Packag.*, pp. 181–184, November.

[106] D. A. Hill, K. Cavcey, and R. T. Johnk, "On the use of the hilbert transform for processing measured cw data," *IEEE Trans. Electromagn. Compat.*, vol. 36, pp. 314–321, November 1994.

[107] C. R. Paul, "Solution of the transmission-line equations under the weak-coupling assumption," *IEEE Trans. Electromagn. Compat.*, vol. 44, pp. 413–423, August 2002.

[108] S. N. Lalgudi, E. Engin, G. Casinovi, and M. Swaminathan, "Accurate transient simulation of interconnects characterized by band-limited data with propagation delay enforcement in a modified nodal analysis framework," *Submitted for review in IEEE Trans. Electromagn. Compat.*, vol. x, pp. xx–xx, July 2007.

[109] D. Winklestein, M. B. Steer, and R. Pomerleau, "Simulation of arbitrary transmission line networks with nonlinear terminations," *IEEE Trans. on Circuits and Systems*, vol. 38, pp. 418–422, April 1991.

[110] W. R. Eisdenstadt, "S-parameter-based ic interconnect transmission line characterization," *IEEE Trans. on Components, Hybrids, and Manufacturing Technology*, vol. 15, pp. 483–490, August 1992.