Title
Statistical Analysis for On-Chip Power Grid Networks and Interconnects Considering Process
Variation

Permalink
https://escholarship.org/uc/item/9n00v9vm

Author
Mi, Ning

Publication Date
2009

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Statistical Analysis for On-Chip Power Grid Networks and Interconnects
Considering Process Variation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Ning Mi

December 2009

Dissertation Committee:
        Dr. Sheldon X.-D. Tan, Chairperson
        Dr. Afshin Abdollahi
        Dr. Frank Vahid

The Dissertation of Ning Mi is approved:

_____

_____

_____
Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First, I would like to express gratefulness to my advisor Dr. Sheldon Tan for his support to my Ph.D study. It is him who bring me into the Electronic Design Automation(EDA) area, and help me explore the long journey in the area. His paper and his way of thinking always inspire me to get to higher level in my research. Besides my advisor, my gratitude goes to the rest of my thesis committee: Dr. Frank Vahid and Dr. Afshin Abdollahi, who give me good questions and encouragement in my defense.

I appreciate the efforts of my colleague, Pu Liu, Wei Wu, Boyuan Yan and Duo Li for their contribution in my research. The discussion with them always gives me more deep understanding in my research area.

Most important are my parents, Yongju Mi and Ping Ma, who always give me encouragement when I have difficulties. They teach me how to fullfil any goal.

Last, but not the least, the financial support of my advisor, my department and University of California, Riverside is gratefully acknowledged.

ABSTRACT OF THE DISSERTATION

Statistical Analysis for On-Chip Power Grid Networks and Interconnects
Considering Process Variation

by

Ning Mi

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, December 2009
Dr. Sheldon X.-D. Tan, Chairperson

With the aggressive scaling down of semiconductor VLSI devices from 65nm to
45 nm, 32nm, the process induced variability becomes the major design concern.
The fundamental change in VLSI chip design in current and future nodes is that
what has been designed will not agree with the products manufactured due to the
uncertainties in the manufacture processes. Even worse, the variabilities keep growing
as the technology scales down continually. The process induced variations manifest
themselves from wafer to wafer, die-to-die and device to device within a die. Some are
systematic variabilities and some are random variabilities, which have to leave extra
margin for worst case. The Monte Carlo method can come to the rescue by simulating
the probability of the worst case in a random way. However, it is well known this
approach is very time consuming and forbidding slow. It is highly desirable to have
more efficient statistical modeling and simulation techniques and tools to guide the
design in the presence of uncertainties in the nanometer VLSI regime.

In this dissertation, the influence of the variability, such as threshold voltage vari-
ation, interconnect wire height, width variation, on the performance of power grid

delivery networks and signal interconnect circuits, are studied. First we develop a new statistical method, which is based on concept of Hermite polynomial chaos, to analyze power grid voltage drop variations of on-chip power grid networks. The new approach considers both wire variations and sub-threshold leakage current variations, which are modeled as lognormal distribution random variables. We also consider spacial correlation of the leakage variables by applying orthogonal decomposition to map the correlated random variables into independent ones before the analysis. Second, we propose a more efficient statistical analysis approach, StoEKS, in which the extended Krylov subspace method is used to speedup the solution procedure of the variational circuit equations. By using the model order reduction technique, StoEKS partially mitigates circuit-size increasing problem associated with the augmented matrices from the Galerkin-based spectral statistical method. Finally, we propose an efficient method to calculate variational interconnect delay, which is a crucial step in the statistical static timing analysis(SSTA). We apply Numerical quadrature method based on orthogonal polynomial representation (OPR) of statistical variations to derive the non-linear, non-Gaussian analytic interconnect delay models in terms of the interconnect wire width, height variations. It can take in variational parameters in OPR form and outputs the delays computed in OPR form again, which is compatible with existing SSTA methods.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

This section presents the motivations of our work from the following three aspects: the variability and uncertainty of integrated circuit, the on-chip power grid network and the variability influence on it.

1.1.1 Variability and uncertainty on VLSI

With the scaling down of semiconductor and the rapid increased complexity of modern nano-meter technique, the difference between the integrated circuit design and manufacturing becomes larger. The variability and uncertainty have to be seriously considered when the scaling down continued [44]. The reason for this is that the CMOS technology leads to the increasing complexity in the devices and the interconnects connecting to them. Also, there are differences, which are induced in the process of manufacturing, existing between the simulation model and the hardware

implementation. Variance induced in the process should be analysis to narrow or even eliminate the gap in design and manufacturing, as to deleminate performance fluctuations in the circuit behavior.

Variability can be classified in various ways. In the following, some classification standards are introduced to present more information of variability in the integrated circuit design and manufacturing.

From the aspect of source, the variation can be classified into physical variability and environment variability. The integrated circuit consists of linear and non-linear devices which fulfill functions for the whole circuit and the interconnect between them. The physical variability are induced when creating these device, such as the procedure of implantation, oxidation, polysilicon line definition, deposition, etching, chemical-mechanical polishing and etc. The physical variability includes threshold voltage variation, thin film thinkness variation, interconnect wire thickness, wideness variation and etc. On the other hand, the circuit performance is still determined by the environment in which the circuit operates. The environment variations, such as temperature, power supply voltage and noise, have as similar impacts on the circuit behavior as the physical variations.

Based on the location of the variation, the variation can be classified as within die or intra-die variation, wafer-to-wafer variation, and chip-to-chip variations. We can exam the spacial behavior of the variability as to give the joint distribution across the die, wafer or chip.

Some kind of variations are known to be a function of specific design characteristics. With a suitable model considering such variability, the manufacture process can meet the requirement of the circuit performance. This kind of variation is referred

as systematic variability because the models can be used to describe the variation in advance. On the other hand, some physical or environment phenomenon is not well understood, not enough available information is given to construct quantitative models. To take this kind of variability into consideration, worst case analysis is given to test the performance of the circuit, such as creating large enough margin for the worst possible condition than may occur, or using statistical method to get the random characteristic of the performing circuit.

As the scaling down of the semiconductor technology, the process variation introduced above becomes more and more severe. As for the systematic variability, models can be built and analysis can be done correctly. However, there are still random variabilities which have to be solved using efficient methods. In this thesis, some statistical methods are proposed to do analysis in power grid delivery network considering process variation.

1.1.2   On-chip Power Grid Network

On-chip power grid network will be introduced in this part, including definition, modeling, simulation and analysis of power grid network. Also, the process variation influencing on the P/G network is briefly proposed in this part.

Power grid distribution networks distribute power and ground voltages from pad locations to all devices on the chip [52].

Fig 1.1 shows a simple on-chip power grid [44]. Power supply wires for V_{dd} are in an orthogonal grid across the various wiring levels, and are connected to the circuit components at the bottom and to the package power terminals at the top.

Large current fluctuation in the power and ground networks, which is caused by the

External voltage source

Power Grid
Delivery network
connecting to
power source

device

device

Integrated Circuit
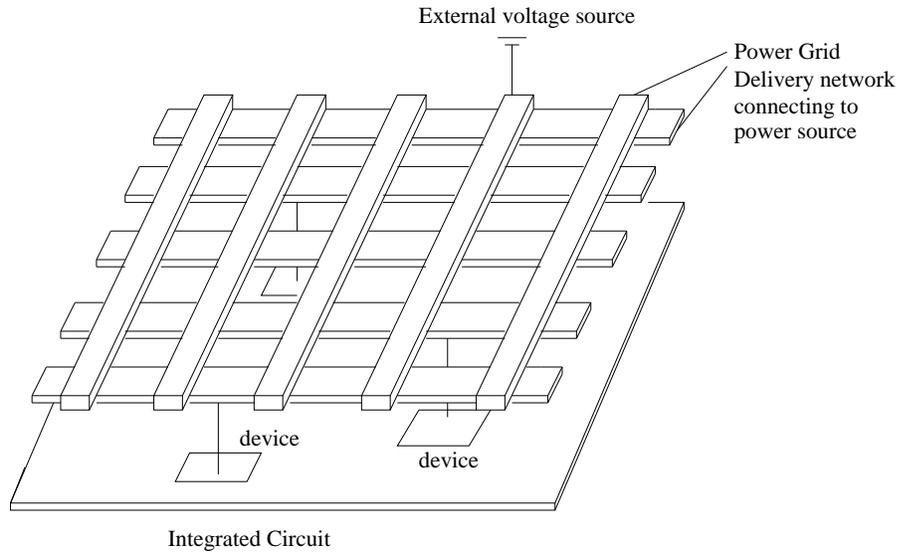
Figure 1.1: Representative of On-Chip Power Grid network

shrinking of the device size, increasing of switching frequencies and increasing power consumption, would degrade the performance and reliability of power distribution networks. Due to the interconnect resistance, there will be voltage drop through the power grid network, which is named as IR drop. Also, the package supplies currents to power grid network either by package lead or by C4 bump array in flip-chip technology. The voltage drop can also be induced here because of the inductance of the package leads, accompanied with the time-varying current drawn by the device on the die, although the resistance of the package is quite small. This kind of voltage drop is referred to as di/dt drop. As a result, the real voltage supplied to the device is the supply voltage minus IR drop and di/dt drop. As the close relationship of circuit delay and power dissipation with power supply voltage [44], the excessive voltage drop on the power grid delivery network would cause excessive delay, which would lead to false behavior of the device. Excessive delay would lead to the slowing down of switching speeds and increasing noise margin of circuit, which would highly

cause function failure. A robust power grid network is vital in meeting performance guarantees and ensuring reliable operation.

Nowadays, tools are used here to do the simulation, analysis, optimization of power grid network. In the process, it has to take into account the power delivery network, on-chip decoupling capacitance, the package parasitics, as well as the parasitics associated with the board and possibly even the power regulator. Usually, to perform power grid network analysis, there are two kinds: (1) DC voltage drop at each component in the circuit and (2) time domain analysis considering capacitive chip and inductive package and time varying current input of power grid network.

Different models are used for power grid delivery network according to different requirements to analysis of the network. DC voltage computation, which is to calculate IR drop, only requires resistance model. Most of the time, when time-varying currents input to network are considered, capacitors should also be considered. The capacitor here includes parasitic wire capacitances, between power wires and ground wire, parasitic capacitance of transistors and explicitly placed decoupling capacitors. When the switching speeding goes high, inductance should also be taken into account because di/dt drop is becoming the important part of the total voltage drop.

It is too complex to simulate the power grid network with all the nonlinear device connecting on the chip at the same time. One of the reasons is the huge size of the network. To put them together to compute is infeasible. Therefore, it takes two steps to do the simulation. First, the nonlinear device is simulated with ideal supply voltages. Then, the currents drawn from the device and connected to the network is computed. Second, with the input current, which is the current of simulated device terminal connecting to power grid network, the voltage drop is simulated in the power

grid network. The reason why the power grid network can be simulated in these two steps is the voltage drop actually is less than %10 of node voltage and therefore, the current interaction between the power grid and device in the circuit can be omitted.

As introduced in chapter 1.1.1, process variation should also be taken into consideration in the analysis of power grid network. The variability includes the current input variability brought from device on the chip, the interconnect variability, such as interconnect wire length and width variability. Chapter4 and chapter5 will introduce methods to analysis power grid delivery network with process variation. Chapter6 will introduce sampling method to analysis interconnect variability for Statistical Static Timing Analysis(SSTA).

1.2   Contributions and Organization

In this thesis, variational on-chip power grid delivery network and variational interconnect delay are studied. Fast analysis approaches are presented here and results are compared with Monte Carlo method.

The rest of this dissertation is organized as follows. Chapter 2 provides background knowledge about on-chip power grid network and interconnect. It includes some models which we will use in latter part of the disertation. Chapter 3 presents knowledge about spectral statistical theory and relative algorithm used in the following chapter. Chapter 4 presents an algorithm to do analysis of on-chip power grid analysis considering leakage current variations. Chapter 5 proposes a fast analysis algorithm to power grid network with consideration of both interconnect variation and input current variation. Chapter 6 introduce a method to calculate variational

interconnect delay for statistical timing analysis. Finally, Chapter 7 concludes the dissertation.

Chapter 2

Background and Related Models

2.1 Variational Power Grid Models

2.1.1 On-chip Power Grid Network Models

The power grid networks in this paper are modeled as RC networks with known time-variant current sources, which are obtained by gate-level logic simulations of the VLSI systems. For a power grid (versus the ground grid), some nodes have known voltages modeled as constant voltage sources. For C4 power grids, the known voltage nodes can be nodes inside the power grid. Given known deterministic vector of current sources, $u(t)$, the node voltages can be obtained by solving the following linear differential equation, which is formulated using modified nodal analysis (MNA) approach,

$$Gv(t) + C\frac{dv(t)}{dt} = Bu(t) \tag{2.1}$$

where $G \in R^{m \times m}$ is the conductance matrix, $C \in R^{m \times m}$ is the admittance matrix resulting from storage elements. $B \in R^{m \times l}$ and $u(t) \in R^{l \times 1}$ is a vector of time-varying node currents. m is the size of the given circuit and l is the number of input ports.

2.1.2 Power Grid Models Considered Process Variations

The G and C matrices and input currents $u(t)$ depend on the circuit parameters, such as metal wire width, length, thickness on power grids, and transistor parameters, such as channel length, width, gate oxide thickness, etc. Some previous work assumes that all circuit parameters and current sources are treated as uncorrelated Gaussian random variables [23]. In this thesis, we consider both power grid wire variations and the log-normal leakage current variations, caused by the channel length variations, which is modeled as Gaussian (normal) variations [50].

Process-induced variations can also be classified into inter-die (die-to-die) variations and intra-die variations. In inter-die variations, all the parameters variations are correlated. The worst case corner can be easily found by setting the parameters to their range limits (mean plus 3σ). The difficulty lies in the intra-die variations, where the circuit parameters are not correlated or spatially correlated within a die. Intra-die variations also consist of local and layout dependent deterministic components and random components, which typically are modeled as multivariate Gaussian process with some spatial correlations [10]. For the spatial correlation as introduced in detail in chapter4, we first assume we have a number of independent (uncorrelated) transformed ortho-normal random Gaussian variables $\xi(\theta), i = 1, ..., n$, which actually model the channel length and the device threshold voltage variations and other variations. Then, we focus on strongly spatial correlation in the intra-die variation.

9

The reasons is that weakly or uncorrelated random variations have smaller impacts on the leakage and wire variations than that of the highly correlated random variables, as the variance of the sum of n independent random variables is $\sim n\sigma^2$, while the variance of the sum of n highly correlated random variables is $\sim n^2\sigma^2$. We apply the principal component analysis (PCA) method to transfer the correlated variables into un-correlated variables before the spectral statistical analysis.

Let Ω denote the sample space of the experimental or manufacturing outcomes. For $\omega \in \Omega$, let $\xi_{\mathbf{d}}(\omega) = [\xi_{1d}(\omega), ..., \xi_{rd}(\omega)]$ be a vector of r Gaussian variables to represent the circuit parameters of interest. After the PCA operation, we obtain independent random variable vectors $\xi = [\xi_1, ..., \xi_n]$. Notice that $n \leq r$ in general. Therefore, given the process variations, the MNA for Eq.2.1 becomes

$$G(\xi)v(t) + C(\xi)\frac{dv(t)}{dt} = Bu(t, \xi(\theta)) \qquad (2.2)$$

The variation in wire width and thickness will cause variation in the conductance matrix $G(\xi)$ and capacitance matrix $C(\xi)$. The variations are more related to back-end-of-the-line (BEOL) as power grids are mainly metals at top or middle layers. The input current vector, $u(t, \xi(\theta))$, has both deterministic and random components. To simplify our analysis, we assume the dynamic currents (power) caused by circuit switching are still modeled as deterministic currents as we only consider the leakage variations. Practically, the variations caused by the dynamic power of circuits can be significant. But the voltage variations caused by the leakage variations can be viewed as background noise, which can be considered together with dynamic power-induced variations later.

To obtain the variation current sources $u(t, \xi(\theta))$, some library characterization methods will be used to compute the $u(t, \xi(\theta))$ once we know the effective channel length (L_{eff}) variations, threshold voltage (V_{th}) variations and other variable sources under different input patterns. With those variation-aware cell library, we can more accurately obtain the $u(t, \xi(\theta))$ based on the logic simulation of the whole chip under some inputs.

Note that practical use perspective, user may be only interested in voltage variations over a period of time or worst case in a period of time. Those information can be easily obtained once we know the variations in any given time instance. In other words, the information we obtain here can be used to derive any other information that is interesting to designers.

The problem we need to solve is to efficiently find the mean and variances of voltage $v(t)$ at any node and at any time instance. A straightforward method is Monte Carlo (MC) based sampling methods. We randomly generate $G(\xi)$, $C(\xi)$ and $u(t, \xi(\theta))$, which is based on the log-normal distribution, solve Eq.2.2 in time domain for each sampling and compute the means and variances based on sufficient samplings. Obviously, MC will be computationally expensive. However, MC will give the most reliable results and is the most robust and flexible method. In this thesis, a method based on spectral statistical analysis is presented, which transforms the statistical differential equation to a series of deterministic equation. It is more efficient than MC. Chapter3 introduce basic knowledge of spectral statistical analysis. Chapter4 and chapter5 presents the methods based on spetral statistical analysis to the statistical analysis of power grid network.

2.2 Variational Interconnect Modeling

2.2.1 Interconnect Modeling

The interconnects here are modeled as RCL linear time-invariant networks. Given known vector of excitation, $u(t)$, the output $x(t)$ can be obtained by solving the following linear differential equation, which is formulated using modified nodal analysis (MNA) approach,

$$Gx(t) + C\frac{dx(t)}{dt} = Bu$$
$$y(t) = L^T x(t) \tag{2.3}$$

with conductance matrix $G \in \Re^{m \times m}$, capacitor matrix $C \in \Re^{m \times m}$, input position matrix $B \in \Re^{m \times l}$ and output matrix $L \in \Re^{m \times l}$. Given the specific input (step, or ramp), we can then compute the delay of the interconnects from the transient waveforms (by measuring the 50%-50% delay or other delay formula).

2.2.2 Interconnect Modeling Considered Process Variations

As introduced in previous, the G and C matrices depend on the circuit parameters, such as metal wire width, length, metal thickness and interlevel-dielectric (ILD) thickness on interconnects. In this thesis, we assume that we have a number of independent (uncorrelated) transformed orth-normal Gaussian random variables $\xi_i(\theta), i = 1, ..., n$, which model the interconnect geometrical variations. The spatial correlation in the intra-die variation can be processed by using the principal component analysis method (or other methods like Karhunen-Loeve transformation and principal factor analysis

etc.) to transform the correlated variables into un-correlated variables before spectral statistical analysis [23, 38]. Let Θ denote the process sampling space. Let $\theta \in \Theta$, $\xi_i : \theta \to R$ denote a normalized Gaussian variable and $\xi(\theta) = [\xi_{1d}(\theta), ..., \xi_{nd}(\theta)]$ is a vector of n Gaussian variable. After orthogonal transformation operation, we obtain independent random variable vectors $\xi = [\xi_1, ..., \xi_n]$. Notice that $n \leq r$ in general. Therefore, given the process variations, the MNA equation for Eq.2.3 becomes

$$G(\xi)x(t,\xi) + C(\xi)\frac{dx(t,\xi)}{dt} = Bu$$
$$y(t,\xi) = L^T x(t,\xi) \tag{2.4}$$

Here, $L = B$. The problem we need to solve is to efficiently compute the delays from the specific input to specific outputs in terms of orthogonal polynomial form under step or ramp inputs without using the time-consuming sampling-based method, such as Monte-Carlo.

Chapter 3

Spectral Statistical Based Simulation

In this chapter, main methods to solve statisitcal differential equation will be reviewed in section 3.1. In section 3.2, the definite and basic concept of Hermite Polynomial Choas will be introduced. Section 3.3 would introduce Galerkin Method, which is a method to solve stochastic differential equation.

3.1   Preview

From chapter 2, we know that the statistical problem needs to solve is in fact a statistical differential equation problem. The system[21] is with random operator and random input as in Eq.3.1:

$$Ax = f \tag{3.1}$$

where A is a statistical differential operator, x is the random response, and f is the possibly random input. Deterministic A and random input f has been studied extensively, while only approximate solutions to the problem are given with both random operator A and random input f.

From the aspect of engineering to solve the statistical differential equation, which would consider with intricate geometries, boundary conditions and various types of excitations, the widely used method is the perturbation method[21]. The reason for this is probably the simplicity of the perturbation method from the engineering aspect. This method is to do expansion to all the random quantities around their respective mean values via a Taylor series. However, it is complex to compute higher order beyond the first and the second order. Also, the instability problem of the approximate solution would appear in higher order terms. The above two lead to the fact that perturbation method also only restricts to involving small randomness.

Another method is the hierarchy closure approximation[7]. This method is based on expressing joint statistical moments of the output and of the system as functions of lower order moments. However, the difficulties is to form higher order moments which limits the method only to small random fluctuations.

Stochastic Green's function[3] and decomposition method[4] for solving nonlinear differential equation are introduced later. The inverse of the random solution process is expanded in a Neumann series[14] or in delta method[54]. However, the high order moments computation is cumbersome since it involves averaging of products of random matrix. The limit here is still the small random fluctuations. Later Monte Carlo Expansion(MCE)[64] method, which constitutes of the Neumann expansion with Monte Carlo simulation is introduced. This method generates random matrix

based on Neumann expansion of inverse of stochastic coefficient matrix. Good match with standard Monte Carlo can be reached with large coefficients of variation. However, the simulation process takes time especially when sizable random fluctuations are involved.

The previous existing solutions to statistical differential equations either limits to the small random fluctuation or to the time consuming problem. The spectral statistical method introduced in the following can be applied to large fluctuation comparing and is much more efficiently than Monte Carlo.

3.2 Stochastic Finite Element

3.2.1 Mathematical Model

In this section, the concept of Hilbert space and spectral analysis will be introduced. Hilbert space of functions defined over a domain $\mathbf{D}$, with value on the real line, is denoted by $\mathbf{H}$. Let$(\Omega, \Gamma, \mathbf{P})$ represents a probability space. Suppose $x \in D$ and $\theta \in \Theta$. Then, the space of functions mapping Ω onto the real line is denoted by Θ. Then Each map $\Theta \to R$ represents a random variable. The equation below represents the random physical model.

$$A(x,\theta)u(x,\theta) = f(x,\theta) \tag{3.2}$$

$A(x,\theta)$ represents random differential operation. In other words, $A(x,\theta)$ is a differential operator with coefficients exhibition random fluctuations with one or more random variables. Here, $u(x,\theta)$ is the function needs to be solved. Usually, as prac-

tical use, the second order expansion of $A(x, \theta)$ is enough because most physical measurement are of the second order type.

The procedure of solving Equation 3.2 includes two steps. First is to use a random distribution to represent $A(x, \theta)$ and $u(x, \theta)$. After obtaining the distribution, the problem turns to how to solve the random equation. Usually, invoking central limit theorem, Gaussian distribution most likely appears in the physical models [21].

As for the solution procedure, usually, a general form of the solution process can be expressed as

$$u = g[\delta(x, \theta), f(x, \theta), x] \tag{3.3}$$

Here, $g[.]$ is some nonlinear functional of its arguments. $\delta(x, \theta)$ is represented as

$$a_i(x, \theta) = \bar{a}_i(x) + \delta(x, \theta) \tag{3.4}$$

$a_i(x, \theta)$, which are the coefficients of $A(x, \theta)$, can be decomposed to deterministic part $\bar{a}_i(x)$ and purely random part $\delta(x, \theta)$. $\bar{a}_i(x)$ is the expectation of $a_i(x)$ and $\delta_i(x, \theta)$ is zero-mean random process.

As in equation 3.3, complete description of the response should includes the joint distribution in it. However, it is out of the capability to compute given the infinite dimensional structure [21]. Finite dimension is required here. The finite dimension representation cannot be achieved through partitioning the space because of the nature of the random process. Therefore, alternatively, an abstraction of the discretization process can be introduced which is mathematically equivalent to a discretization with respect to a spectral measure.

The following will introduce the method of spectral representation of random

process which is in the use of following work.

There are two kinds of representations used in the development of the stochastic finite element method. They are Karhunen-Loeve expansions and the Homogeneous Chaos(PC) expansion. The Karhunen-Loeve expansion method requires knowing the covariance function of the process that needs to be expanded. It is able to be obtained for the operation part. However, it cannot be implemented for the solution process, since the covariance function of solution part is always not known.

The Homogeneous Choas expansion involves a basis of known random functions with deterministic coefficients. This expansion also means to be found by minimizing some norm of the error resulting from a finite representation. Equation 3.3 is rewritten as

$$u = h[\xi_i(\theta), x] \tag{3.5}$$

Here, ξ_i represents the random variables in the random system. If the variables are following Gaussian distribution, which means the nonlinear expansion of $h[.]$ are in terms of Gaussian, the expansion is known as Homogenous Choas. In other words, it can be viewed as an orthogonal development for nonlinear functionals with Gaussian measure.

3.2.2 Concept of Hermite PC

In the following, a random variable $\xi(\theta)$ is expressed as a function of θ, which is the random event. Hermite PC utilizes a series of orthogonal polynomials (with respect to the Gaussian distribution) to facilitate stochastic analysis [63]. These polynomials are used as the orthogonal base to decompose a random process in a similar way that

sine and cosine functions are used to decompose a periodic signal in a Fourier series expansion.

Note that for the Gaussian and log-normal distributions, Hermite polynomial is the best choice as they lead to exponential convergence rate [21]. For non Gaussian and non log-normal distributions, there are other orthogonal polynomials such as Legendre for uniform distribution, Charlier for Poisson distribution and Krawtchouk for Binomial distribution etc [20, 59].

For a random variable $v(t, \xi)$ with limited variance, where $\xi = [\xi_1, \xi_2, ...\xi_n]$ is a vector of zero mean orthonormal Gaussian random variables. The random variable can be approximated by truncated Hermite PC expansion as follows [21]:

$$v(t, \xi) = \sum_{k=0}^{P} a_k H_k^n(\xi) \tag{3.6}$$

where n is the number of independent random variables, $H_k^n(\xi)$ is n-dimensional Hermite polynomials and a_k are the deterministic coefficients. The number of terms P is given as

$$P = \sum_{k=0}^{p} \frac{(n-1+k)!}{k!(n-1)!} \tag{3.7}$$

where p is the order of the Hermite PC. If only one random variable is considered, the one-dimensional Hermite polynomials are expressed as follows:

$$H_0^1(\xi) = 1, H_1^1(\xi) = \xi, H_2^1(\xi) = \xi^2 - 1, H_3^1(\xi) = \xi^3 - 3\xi, ... \tag{3.8}$$

Hermite polynomials are orthogonal with respect to Gaussian weighted expectation

(the superscript n is dropped for simple notation):

$$< H_i(\xi), H_j(\xi) >=< H_i^2(\xi) > \delta_{ij} \tag{3.9}$$

where δ_{ij} is the Kronecker delta and $< *, * >$ denotes an inner product defined as follow:

$$< f(\xi), g(\xi) >= \frac{1}{\sqrt{(2\pi)^n}} \int f(\xi)g(\xi)e^{-\frac{1}{2}\xi^T\xi}d\xi \tag{3.10}$$

Like Fourier series, the coefficient a_k can be found by a projection operation onto the Hermite PC basis:

$$a_k(t) = \frac{< v(t, \xi), H_k(\xi) >}{< H_k^2(\xi) >}, \ \forall k \in \{0, ..., P\}. \tag{3.11}$$

3.3   Simulation Approach Based on Hermite PCs

To simplify the presentation, we first assume that C and G are deterministic in Eq.2.2. We will remove this assumption later. In case that $v(t, \xi)$ is unknown random variable vector (with unknown distributions) like node voltages in Eq.2.2, then the coefficients can be computed by using Galerkin method, which states that the best approximation of $v(t, \xi)$ is obtained when the error $\Delta(t, \xi)$, which is defined as

$$\Delta(t, \xi) = Gv(t) + C\frac{dv(t)}{dt} - Bu(t, \xi(\theta)) \tag{3.12}$$

is orthogonal to the Hermite polynomials. That is

$$< \Delta(t, \xi), H_k(\xi) >= 0, \ i = 0, 1, ..., P \tag{3.13}$$

20

In this way, we transform the statistical analysis process to a deterministic process, where we only need to compute the coefficients of its Hermite PC. Once we obtain those coefficients, the mean and variance of the random variables can be easily computed.

Chapter 4

Statistical Analysis considering Leakage Current Variations With Spatial Correlation

4.1 Problem Statement

4.1.1 Leakage Current in Power Grid Network

One important aspect of the variations comes from the chip leakage currents. Leakage currents come from different sources. The dominant factor is the sub-threshold leakage current. The reason is that sub-threshold leakage current has a rapid increasing rate (about 5X-10X increase per technology generation [16]), and it is highly sensitive to threshold voltage V_{th} variations, owning to the exponential relationship between

sub-threshold current I_{off} and threshold voltage V_{th} as shown below [56],

$$I_{off} = I_{s0}e^{\frac{V_{gs}-V_{th}}{nV_T}}\left(1 - e^{\frac{-V_{ds}}{V_T}}\right) \qquad (4.1)$$

where I_{s0} is a constant related to the device characteristics, V_T is the thermal voltage, and n is a constant. It was shown in [28] that leakage variations for 90nm can be 20×. Based on the ITRS 2005 [2], the leakage power accounts for more than 60% at 45nm, there are many consequences for chip design, especially for design of the power grid. The grid will develop voltage drop at all the nodes that are correspondingly significant with a strong within-die components. The voltage drop is unavoidable and manifests itself as a background noise on the grid which has an impact on the circuit delay and operation.

Clearly, the leakage current has exponential dependency on the threshold voltage V_{th}. In the sequel, the leakage current is mainly referred to as the sub-threshold leakage current. Detailed analysis shows that I_{off} is also an exponential function of the effective channel length L_{eff} [50]. Actually L_{eff} are strongly correlated with V_{off} as V_{off} variations typically are caused by the L_{eff}. So if we model V_{th} or L_{eff} as the random variables with Gaussian variation caused by the inter-die or intra-die process variations, then the leakage currents will have a log-normal distribution as shown in [50]. On top of this, those random variables are spatially correlated within a die, owning to the nature of the many physical and chemical manufacture processes [40].

4.1.2 Statistical analysis of Power Grid with Leakage Variations

On-chip power grid analysis and designs have been intensively studied in the past due to the increasing impacts of excessive voltage drops as technologies scale [61, 65, 29, 51, 57, 58, 13, 24, 62, 17, 53]. Owning to the increasing impacts of leakage currents and its variations on the circuit performances, especially on the on-chip power delivery networks, a number of research works have been proposed recently to perform the statistical analysis of power grid networks under process-induced leakage current variations. The voltage drop of power grid networks subject to the leakage current variations was first studied in [18, 19]. This method assumes that the log-normal distribution of the node voltage drop caused by the log-normal leakage current inputs and is based on a localized Monte Carlo (sampling) method to compute the variance of the node voltage drop. However, this localized sampling method is limited to the static DC solution of power grids modeled as resistor-only networks. Therefore, it can only compute the responses to the standby leakage currents. However, the dynamic leakage currents become more significant, especially when the sleep transistors are intensively used nowadays for reducing leakage powers. In [55, 45], impulse responses are used to compute the means and variances of node voltage responses caused by general current variations. But this method needs to know the impulse response from all the current sources to all the nodes, which is expensive to compute for a large network. In [50], the probability density function (pdf) of leakage currents is computed based on the Gaussian variations of channel lengths.

Recently, a statistical simulation method for interconnect and power grid networks has been proposed [60, 23, 59]. This method is based on the orthogonal polynomial

chaos expansion of random processes to represent and solve for the statistical responses of linear systems. The major benefit of this method is its compatibility with current transient simulation framework: it solves for some coefficients of the orthogonal polynomials, which can be done by using normal transient simulations of the original circuits with deterministic inputs to compute variances of node responses. Some existing approaches [23, 59] model all the parameter variations as Gaussian (or approximate them as Gaussian variations by using first-order Taylor expansion) [60]. Those methods also fail to consider the spatial correlation in the process parameter random variables.

In this chapter, we apply the orthogonal polynomial based method (also called spectral statistical method) to deal with leakage current inputs with log-normal distributions and spatial correlations. We show how to represent a log-normal distribution in terms of Hermite polynomials, assuming Gaussian distribution of threshold voltage V_{th} in consideration of intra-die variation. To consider the spatial correlation, we apply orthogonal decomposition via principal component analysis to map the correlated random variables into independent variables. To the best knowledge of the authors, the proposed method is the first method being able to perform statistical analysis on power grids with variation dynamic leakage currents having lognormal distributions and spatial correlations. Experiment results show that the proposed method predicates the variances of the resulting log-normal-like node voltage drops more accurately than Taylor expansion based Gaussian approximation method.

We remark that we only consider the leakage current inputs with log-normal distributions to emphasize our new contributions. The reason is that leakage currents can be variable significantly. In 90nm, it can lead to 20X variations [28]. For the

coming 45nm, it will dominate the currents of a chip (60% based on ITRS 2005 [2]). Considering variations from leakage currents has significant practice relevance. Also for general current variations from dynamic power of the circuits, which typically can be modeled as Gaussian distribution, existing work [23] using Taylor series expansion has been explored. The voltage variations caused by the dynamic power can be considered on top of the variations from the log-normal leakage currents. We notice that similar work, which considers only leakage variations have been done before [18, 19].

Also we remark that Vdd drop will have impacts on the leakage currents, which create a negative feedback for the leakage current itself as increasing Vdd drop leads to lower V_{gs} in Eq.4.1, which leads to smaller I_{off}. However, to consider the effect, both the power grid and signal circuits need to be simulated together, which will be very expensive. Hence practically, two-step simulation approach is used where power grid and signal circuits are simulated separately but in an iterative way to consider the coupling between them. In light of this simulation methodology, our proposed method can be viewed as the only one step (power grid simulation step) in such a method.

4.2 Hermite PCs For Log-Normal Leakage Current Variation

In this section, we present the new method for representing the log-normal leakage current distributions by using Hermite PCs independent Gaussian variables representing the channel length or threshold voltage variations. Our method, based on [20], can be applied to one or more independent Gaussian variables.

4.2.1 Using Galerkin equation for independent random variable

For illustration purpose, considering one Gaussian variable $\xi = [\xi_1]$ and we then can assume that the node voltage response can be written as a second order $(p = 2)$ Hermite PC:

$$v(t, \xi) = v_0(t) + v_1(t)\xi_1 + v_2(t)(\xi_1^2 - 1) \tag{4.2}$$

Assuming that the input leakage current sources can also be represented by a second Hermite PC:

$$u(t, \xi) = u_0(t) + u_1(t)\xi_1 + u_2(t)(\xi_1^2 - 1) \tag{4.3}$$

By applying the Galerkin equation Eq.3.13 and note the orthogonal property of the various orders of Hermite PCs, we end up with the following equations

$$Gv_i(t) + C\frac{dv_i(t)}{dt} = Bu_i(t) \tag{4.4}$$

where $i = 0, 1, 2, .., P - 1$.

For two independent Gaussian variables, we have

$$
\begin{aligned}
v(t, \xi) &= v_0(t) + v_1(t)\xi_1 + v_2(t)\xi_2 + v_3(t)(\xi_1^2 - 1) \\
&\quad + v_4(t)(\xi_2^2 - 1) + v_5(\xi_1\xi_2)
\end{aligned} \tag{4.5}
$$

Assuming that we have a similar second order Hermite PC for input leakage current

$u(t, \xi)$,

$$
\begin{aligned}
u(t, \xi) &= u_0(t) + u_1(t)\xi_1 + u_2(t)\xi_2 + u_3(t)(\xi_1^2 - 1) \\
&\quad + u_4(t)(\xi_2^2 - 1) + u_5(\xi_1\xi_2)
\end{aligned} \tag{4.6}
$$

The Eq.4.4 is valid with $i = 0, ..., 5$. For more (more than two) Gaussian variables, we can obtained the similar results with more coefficients of Hermite PCs to be solved by using Eq.4.4.

Once we obtain the Hermite PC of $v(t, \xi)$, we can obtain the mean and variance of $v(t, \xi)$ trivially as (one Gaussian variable case):

$$
\begin{aligned}
E(v(t, \xi)) &= v_0(t) \\
Var(v(t, \xi)) &= v_1^2(t)Var(\xi_1) + v_2^2(t)Var(\xi_1^2 - 1) \\
&= v_1^2(t) + 2v_2^2(t)
\end{aligned} \tag{4.7}
$$

One critical problem remains so far is how to obtain the Hermite PC Eq.4.3 for leakage current with log-normal distribution.

4.2.2 Hermite PC representation of log-normal variables

Let $g(\xi)$ be the Gaussian random variable, denoting threshold voltage or device channel length. Let $l(\xi)$ be the random variable obtained by taking the exponential of $g(\xi)$

$$
l(\xi) = e^{g(\xi)}, \quad g(\xi) = ln(l(\xi)) \tag{4.8}
$$

Obviously, for the MOS device leakage current equation Eq.4.1, leakage current, $I_{off} = cI_l(V_{th}) = ce^{-V_{th}}$, where the leakage component $I_l(V_{th})$ is a log-normal random variable. Let the mean and the variance of $g(\xi)$ as μ_g and σ_g^2, then the mean and variance of $l(\xi)$ are

$$\mu_l = e^{(\mu_g + \frac{\sigma_g^2}{2})} \tag{4.9}$$

$$\sigma_l^2 = e^{(2\mu_g + \sigma_g^2)}[e^{\sigma_g^2} - 1] \tag{4.10}$$

respectively.

For a general Gaussian variable $g(\xi)$, it can always be represented in the following affine form:

$$g(t, \xi) = \sum_{i=0}^{n} \xi_i g_i(t) \tag{4.11}$$

where ξ_i are orthonormal Gaussian variables. i.e. $< \xi_i \xi_j > = \delta_{ij}$, $< \xi_i > = 0$ and $\xi_0 = 1$ and g_i is the coefficient of the individual Gaussian variables. Note that such form can always be obtained by using Karhunen-Loeve orthogonal expansion method [21]

In our problem, we need to represent the log-normal random variable $l(t, \xi)$ by using the Hermite PC expansion form:

$$l(t, \xi) = \sum_{k=0}^{P-1} l_k(t) H_k^n(\xi) \tag{4.12}$$

where $l_0(t) = \exp[\mu_g(t) + \frac{\sigma_g^2}{2}]$. Here the mean is timing varying and standard deviation is not changed with time. To find the other coefficients, we can apply Eq.3.11 on

$l(t, \xi)$. Therefore, we have

$$l_k(t) = \frac{< l(t, \xi), H_k(\xi) >}{< H_k^2(\xi) >}, \quad \forall k \in \{0, ..., P\}. \tag{4.13}$$

It was shown in [20], $l(t, \xi)$ can be written as

$$l(t, \xi) = \frac{< H_k(\xi - \mathbf{g}) >}{< H_i^2(\xi) >} = \exp[\mu_g + \frac{1}{2} \sum_{j=1}^{n} g_j^2] \tag{4.14}$$

where n is the number of independent Gaussian random variables.

The log-normal process can then be written as

$$u(t, \xi) = l_0(t)(1 + \sum_{i=1}^{n} \xi_i g_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\xi_i \xi_j - \delta_{ij})}{< (\xi_i \xi_j - \delta_{ij})^2 >} g_i g_j + ...) \tag{4.15}$$

where g_i is defined in Eq.4.11.

4.2.3 Hermite PC representation with one Gaussian variable

In this case, $\xi = [\xi_1]$. For the second order Hermite PC ($P = 2$), following Eq.4.15, we have

$$u(t, \xi) = l_0(t)(1 + \sigma_g \xi_1 + \frac{1}{2}\sigma_g^2(\xi_1^2 - 1)) \tag{4.16}$$

Hence, the desired Hermite PC coefficients, $u_{0,1,2}$, can be expressed as $l_0, l_0\sigma_g$ and $\frac{1}{2}l_0\sigma_g^2$ respectively.

4.2.4 Hermite PC representation of two and more Gaussian variables

For two random variables ($n = 2$), assume that $\xi = [\xi_1, \xi_2]$ is a normalized uncorrelated Gaussian random variable vector that represents random variable $g(\xi)$:

$$g(t, \xi) = \mu_g(t) + \sigma_1 \xi_1 + \sigma_2 \xi_2 \qquad (4.17)$$

Note that

$$< (\xi_i \xi_j - \delta_{ij})^2 > = < \xi_i^2 \xi_j^2 > = < \xi_i^2 > < \xi_j^2 > = 1$$

Therefore, the expansion of the log-normal random variables using second order Hermite PCs can be expressed as

$$
\begin{aligned}
u(t, \xi) &= l_0(t)(1 + \sigma_1 \xi_1 + \sigma_2 \xi_2 + \frac{\sigma_1^2}{2}(\xi_1^2 - 1) + \frac{\sigma_2^2}{2}(\xi_2^2 - 1) + \\
& \quad 2\sigma_1 \sigma_2 \xi_1 \xi_2)
\end{aligned}
\qquad (4.18)
$$

where

$$\mu_l(t) = l_0(t) = \exp(\mu_g(t) + \frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2)$$

Hence, the desired Hermite PC coefficients, $u_{0,1,2,3,4,5}$, can be expressed as l_0, $l_0\sigma_1$, $l_0\sigma_2$, $\frac{1}{2}l_0\sigma_1^2$, $\frac{1}{2}l_0\sigma_2^2$, and $2l_0\sigma_1\sigma_2$ respectively.

Similarly, for four Gaussian random variables, assume that $\xi = [\xi_1, \xi_2, \xi_3, \xi_4]$ is a normalized, uncorrelated Gaussian random variable vector. The random variable $g(t, \xi)$ can be expressed as

$$g(t, \xi) = \mu_{\mathbf{g}}(\mathbf{t}) + \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{4}} \sigma_{\mathbf{i}} \xi_{\mathbf{i}} \qquad (4.19)$$

As a result, the log-normal random variable $l(\xi)$ can be expressed as

$$u(t, \xi) = l_0(t)(1 + \sum_{i=1}^{4} \xi_i \sigma_i + \sum_{i=1}^{4} \frac{1}{2}(\xi_i^2 - 1)\sigma_i^2 + \sum_{i=1}^{4}\sum_{j=1}^{4} \xi_i \xi_j \sigma_i \sigma_j + ...) \qquad (4.20)$$

where

$$\mu_l(t) = l_0(t) = \exp(\sigma_0 + \frac{1}{2}\sum_{i=1}^{4} \sigma_i^2)$$

Hence, the desired Hermite PC coefficients can be expressed using the equation Eq.4.20 above.

Once we have the Hermite PC representation of the leakage current sources $u(t, \xi)$, the node voltages $v(t, \xi)$ can be computed by using equations Eq.4.4 with proper order p of the PCs to obtain all the Hermite PC coefficients of $v(t, \xi)$.

4.3 Spatial Correlation

In this section, we consider the spatial correlation among different variations within a die. Spatial correlations exist in the intra-die variations in different forms and have been modeled for timing analysis [41, 10]. The general way to consider spatial correlation is by means of mapping the correlated random variables into a set of independent variables. This can be done by using some orthogonal mapping techniques, such as principal component analysis(PCA). In this part, we also apply PCA method in our spectral statistical analysis framework for power/grid statistical analysis.

4.3.1 Concept of principal component analysis

We first briefly review the concept of principal component analysis, which is used here to transform the random variables with correlation to uncorrelated random variables [26].

Suppose that x is a vector of n random variables, $x = [x_1, x_2, ..., x_n]^T$, with covariance matrix C and mean vector $\mu_x = [\mu_{x_1}, \mu_{x_2}, ..., \mu_{x_n}]$. To find the orthogonal random variables, we first calculate the eigenvalue and corresponding eigenvector of covariance matrix C. Then, by ordering the eigenvectors in descending order eigenvalues, the orthogonal matrix A will be obtained. Here, A is expressed as

$$A = [e_1^T, e_2^T, ..., e_n^T]^T \tag{4.21}$$

where e_i is the corresponding eigenvector to eigenvalue λ_i, which satisfies

$$\lambda_i e_i = C e_i, i = 1, 2, ..., n \tag{4.22}$$

and

$$\lambda_i < \lambda_{i-1}, \ i = 2, 3, ..., n \tag{4.23}$$

With A, we can perform the transformation to get orthogonal random variables y, $y = [y_1, y_2, ..., y_n]^T$ by using

$$y = A(x - \mu_x) \tag{4.24}$$

where, y_i is a random variable with Gaussian distribution. The mean, μ_{y_i}, is 0 and

the standard deviation, σ_{y_i}, is $\sqrt{\lambda_i}$ on the condition that [26]

$$e_i^T e_i = 1, i = 1, 2, ..., n \tag{4.25}$$

Here, because of the orthogonal property of matrix A

$$A^{-1} = A^T \tag{4.26}$$

To reconstruct the original random variables, we use the following equation:

$$x = A^T y + \mu_x \tag{4.27}$$

4.3.2　Spatial correlation in statistical power grid analysis

To consider intra-die variation in V_{th}, the chip is divided into n regions. Assuming $\Phi = [\Phi_1, \Phi_2, ..., \Phi_n]$ is a random variable vector, representing the variation of V_{th} on different part of the circuit. In other words, in the **i**th region, the leakage current $I_{off_i} = ce^{V_{th}(\Phi_i)}$, follows the log-normal distribution. Here, Φ_i is a random variable with Gaussian distribution. $\mu_\Phi = [\mu_{\Phi_1}, \mu_{\Phi_2}, ..., \mu_{\Phi_n}]$ is the mean vector of Φ and C is the covariance matrix of Φ.

With PCA, we can get the corresponding uncorrelated random variables $\phi = [\phi_1, \phi_2, ..., \phi_n]$ from the equation

$$\phi = A(\Phi - \mu_\Phi) \tag{4.28}$$

Also, the original random variables can be expressed as

$$\Phi_i = \sum_{j=1}^{n} a_{ij}\phi_j + \mu_{\Phi_i}, i = 1, 2, ...n \tag{4.29}$$

where a_{ij} is the ith row, jth column element in the orthogonal mapping matrix defined in Eq.4.24. $\phi = [\phi_1, \phi_2, ..., \phi_n]$ is a vector with orthogonal Gaussian random variables. The mean of ϕ_j is 0 and variance is λ_j, $j = 1, 2, ..., n$. The distribution of ϕ_i can be written as

$$\phi_i = \mu_{\phi_i} + \sigma_{\phi_i}\hat{\xi}_i, i = 1, 2, ..., n \tag{4.30}$$

$\hat{\xi} = [\hat{\xi}_1, \hat{\xi}_2, ..., \hat{\xi}_n]$ is a vector with orthogonal normal Gaussian random variable. Φ_i can be expressed with normal random variables, $\hat{\xi} = [\hat{\xi}_1, \hat{\xi}_2, ..., \hat{\xi}_n]$:

$$\Phi_i = \sum_{j=1}^{n} a_{ij}\sqrt{\lambda_j}\hat{\xi}_j + \mu_{\Phi_i}, i = 1, 2, ..., n \tag{4.31}$$

With Eq.4.31, the leakage current can be expanded as Hermite Polynomial Chaos:

$$u(\Phi_i) \sim e^{\Phi_i} = e^{\sum_{j=1}^{n} g_j\hat{\xi}_j + \mu_{\Phi_i}} = \mu_i(1 + \sum_{j=1}^{n} \hat{\xi}_j g_j +$$
$$\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{(\hat{\xi}_j\hat{\xi}_k - \delta_{jk})}{< (\hat{\xi}_j\hat{\xi}_k - \delta_{jk})^2 >} g_j g_k + ...) \tag{4.32}$$

Here,

$$g_j = a_{ij}\sqrt{\lambda_j}, j = 1, 2, ..., n \tag{4.33}$$

Therefore, the MNA equation with correlated random variables Φ in current source

35

can be expressed in terms of uncorrelated random variables $\hat{\xi}$ as follows:

$$Gv(t) + C\frac{dv(t)}{dt} = Bu(t, \hat{\xi}) \tag{4.34}$$

With orthogonal property of $\hat{\xi}$, Eq.4.34 will be simply solved by using Eq.4.4, $i = 0, 1, ..., P - 1$.

4.3.3 Variations in wires and leakage currents

In this section, we will consider variations in width (W), thickness(T) of wires of power grids, as well as threshold voltage(V_{th}) in active devices which are reflected in the leakage currents. Meanwhile, without loss of generality, these variations are supposed to be independent of each other. As mentioned in [23], the MNA equation for the ground circuit will becomes:

$$G(\xi_g)v(t, \xi) + C(\xi_c)\frac{dv(t, \xi)}{dt} = Bu(\xi_u, t) \tag{4.35}$$

The variation in width W and thickness T will cause variation in conductance matrix G and capacitance matrix C while variation in threshold voltage will cause leakage current input variation in u. Thus, the conductance and capacitance of wires can be expressed as in [23]:

$$G(\xi_g) = G_0 + G_1\xi_g$$
$$C(\xi_c) = C_0 + C_1\xi_c \tag{4.36}$$

G_0, C_0 represents the deterministic components of conductance and capacitance of the wires. G_1, C_1 represents sensitivity matrices of the conductance and capacitance. ξ_g, ξ_c are normalized random variables with Gaussian distribution, representing process variation in wires of conductance and capacitor, respectively. As mentioned in previous section, the variation in leakage current can be represented by a second Hermite PC as in equation Eq.4.16:

$$u(t, \xi_u) = u_0(t) + u_1(t)\xi_u + I_2(t)(\xi_u^2 - 1) \tag{4.37}$$

here, ξ_u is a normalized Gaussian distribution random variable representing variation in threshold voltage. $u(t, \xi_u)$ follows lognormal distribution as

$$u = e^{g(\xi_u)}$$

$$g(\xi_u) = \mu_u + \sigma_u \xi_u \tag{4.38}$$

As in previous part, the desired Hermite PC coefficients, $u_{0,1,2}$, can be expressed as $u_0, u_0\sigma_u$ and $\frac{1}{2}u_0\sigma_u^2$ respectively. u_0 is the mean of leakage current source, which is expressed as

$$u_0 = e^{\mu_u + \frac{1}{2}\sigma_u^2} \tag{4.39}$$

Considering the influence of ξ_g, ξ_c, ξ_u, the node voltage is therefore expended by Hermite PC in the second order form as

$$
\begin{aligned}
v(t,\xi) =\ & v_0(t) + v_1(t)\xi_g + v_2(t)\xi_c + v_3(t)\xi_u \\
& +v_4(t)(\xi_g^2 - 1) + v_5(t)(\xi_c^2 - 1) + v_6(t)(\xi_u^2 - 1) \\
& +v_7(t)\xi_g\xi_c + v_8(t)\xi_g\xi_u + v_9(t)\xi_c\xi_u
\end{aligned}
\tag{4.40}
$$

Now the task is to compute coefficients of the Hermite PC of node voltage $v(t,\xi)$. Applying Galerkin equation Eq.3.13, we only need to solve the equations as follows:

$$
\begin{aligned}
< \Delta(t,\xi), 1 >= 0; \quad & < \Delta(t,\xi), \xi_g >= 0; \\
< \Delta(t,\xi), \xi_c >= 0; \quad & < \Delta(t,\xi), \xi_u >= 0; \\
< \Delta(t,\xi), \xi_g^2 - 1 >= 0; \quad & < \Delta(t,\xi), \xi_c^2 - 1 >= 0; \\
< \Delta(t,\xi), \xi_u^2 - 1 >= 0; \quad & < \Delta(t,\xi), \xi_g\xi_c >= 0; \\
< \Delta(t,\xi), \xi_g\xi_u >= 0; \quad & < \Delta(t,\xi), \xi_c\xi_u >= 0;
\end{aligned}
\tag{4.41}
$$

With the distribution of $\xi_{\mathbf{g}}, \xi_{\mathbf{c}}, \xi_{\mathbf{u}}$, we can get these coefficients $v(t) = [v_0(t), v_1(t), ..., v_9(t)]^T$ of node voltage as

$$
\widetilde{G}v(t) + \widetilde{C}\frac{dv(t)}{dt} = \widetilde{u}(t)
\tag{4.42}
$$

$$\widetilde{G} \;=\; \begin{bmatrix} G_0 & G_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ G_1 & G_0 & 0 & 0 & 2G_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & G_0 & 0 & 0 & 0 & 0 & G_1 & 0 & 0 \\ 0 & 0 & 0 & G_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & G_1 & 0 & 0 & G_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & G_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & G_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & G_0 & 0 & 0 \\ 0 & 0 & 0 & G_1 & 0 & 0 & 0 & 0 & G_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & G_0 \end{bmatrix}$$

$$(4.43)$$

$$\widetilde{C} \;=\; \begin{bmatrix} C_0 & 0 & C_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & C_0 & 0 & 0 & 0 & 0 & 0 & C_1 & 0 & 0 \\ C_1 & 0 & C_0 & 0 & 0 & 2C_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & C_1 & 0 & 0 & C_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & C_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_0 & 0 \\ 0 & 0 & 0 & C_1 & 0 & 0 & 0 & 0 & 0 & C_0 \end{bmatrix}$$

$$(4.44)$$

$$\widetilde{u}(t) \;\; = \;\; [u_0(t), 0, 0, u_1(t), 0, 0, u_2(t), 0, 0, 0]^T \qquad\qquad (4.45)$$

Knowing Hermite PC coefficients of node voltage $v(t, \xi)$, it is easy to get the mean and variance of $v(t, \xi)$, which describe the random characteristic of node voltage in the given circuit.

We remark that the proposed method will lead to large circuit matrices, which will add more computation costs. To mitigate this scalability problem, for really large power grid circuits, we can apply partitioning strategies to compute the variational responses for each subcircuits, which will be small enough for efficient computation, as done in the existing work [65, 13]. We can also use model order reduction like simulation method, which will be introduced in next Chapter.

4.4   Experiments

This section describes the simulation results of circuits with log-normal leakage current distributions for a number of power grid networks. All the proposed methods have been implemented in Matlab. Sparse techniques are used in the Matlab. All the experimental results are carried out in a Linux system with dual Intel Xeon CPUs with 3.06Ghz and 1GB memory.

The power grid circuits we test are RC mesh circuits based on the values from some industry circuits, which are driven by only leakage currents as we are only interested in the variations from the leakage currents. The resistor values are in the range $10^{-2}\Omega$ and capacitor values are in the range of 10^{-12} farad.

4.4.1 Comparison with Taylor expansion method

We first compare the proposed method with the simple Taylor expansion method for one and more Gaussian variables.

For simplicity, we assume one Gaussian random variable $g(\xi)$, which is expressed as

$$g = \mu_g + \sigma_g \xi \tag{4.46}$$

where ξ is a normalized Gaussian random variable with $< \xi > = 0$, and $< \xi^2 > = 1$.

The log-normal random variable $l(\xi)$, obtained from $g(\xi)$, is written as

$$l(\xi) = e^{g(\xi)} = \exp(\mu_g + \sigma_g \xi) \tag{4.47}$$

Expand the exponential into Taylor series and keep all the terms up to second order, then we have

$$
\begin{aligned}
l(\xi) &= 1 + \sum_{i=0}^{1} \xi_i g_i + \frac{1}{2} \sum_{i=0}^{1} \sum_{j=0}^{1} \xi_i \xi_j g_i g_j + ... \\
&= 1 + \mu_g + \frac{1}{2}\mu_g^2 + \frac{1}{2}\sigma_g^2 + (\sigma_g + \mu_g \sigma_g)\xi + \\
&\quad \frac{1}{2}\sigma_g^2(\xi^2 - 1) + ...
\end{aligned} \tag{4.48}
$$

We observe that the second-order Taylor expansion, as shown in Eq.4.48, is similar to second order Hermite PC in Eq.4.18. Hence, the Galerkin method can still be applied, we then use Eq.4.4 to obtain the Hermite PC coefficients of node voltage $v(t, \xi)$ accordingly. We want to emphasize, however, that the polynomials generated by Taylor expansion in general are not orthogonal with respect to Gaussian distribu-

δ_g	0.01	0.1	0.3	0.5	0.7
HPC (%)	3.19	1.88	2.07	5.5	2.92
Taylor (%)	3.19	1.37	2.41	16.6	24.02

Table 4.1: Accuracy comparison between Hermite PC (HPC) and Taylor Expansion.

tions and can't be used with Galerkin method, unless we only keep the first order of Taylor expansion results (with less accuracy). In this case, the resulting node voltage distribution is still Gaussian, which obviously is not correct.

We note that the first order Taylor expansion has been used in the statistic timing analysis [10]. The delay variations, owning to interconnects and devices, can be approximated with this limitation. The skew distributions may be computed easily with Gaussian distribution.

To compare these two methods, we use the Monte Carlo method to measure the accuracies of two methods in terms of standard deviation. For Monte Carlo, we sample 2000 times, which represents 97.7% accuracy. The results are summarized in Table 4.1. In this table, δ_g is the standard deviation of the Gaussian random threshold voltage Gaussian variable in the log-normal current source, *HPC* is the standard deviation from the Hermite PC method in terms of relative percentage against the MC method. *Taylor* is the standard deviation from the Taylor expansion method in terms of relative percentage against the MC method.

We can observe that when the variation of current source increases, the Taylor expansion method will result in significant errors compared to the MC method. While the proposed method has the smaller errors for all cases. This clearly shows the advantage of the proposed method, which is suitable for big variation.

4.4.2 Examples without spatial correlation

Fig.4.1 shows the node voltage distribution at one node on a certain point of a ground network with 1720 nodes. The Monte Carlo results are obtained by 2000 samples. The standard deviations of the log-normal current sources with one Gaussian variable is 0.1. The mean and 3σ computed by the Hermite PC method are also marked in the figure which fits very well with the MC results. Fig.4.2 shows the node voltages and its variations caused by the leakage currents from 0ns to 126ns. The circuit selected contains 64 nodes with one Gaussian variable of 0.06 in the current source. The blue dotted lines are mean, upper bound and lower bound. The cyan lines are node voltages of Monte Carlo with 2000 times. Most of the MC results are in between upper bound and lower bound.
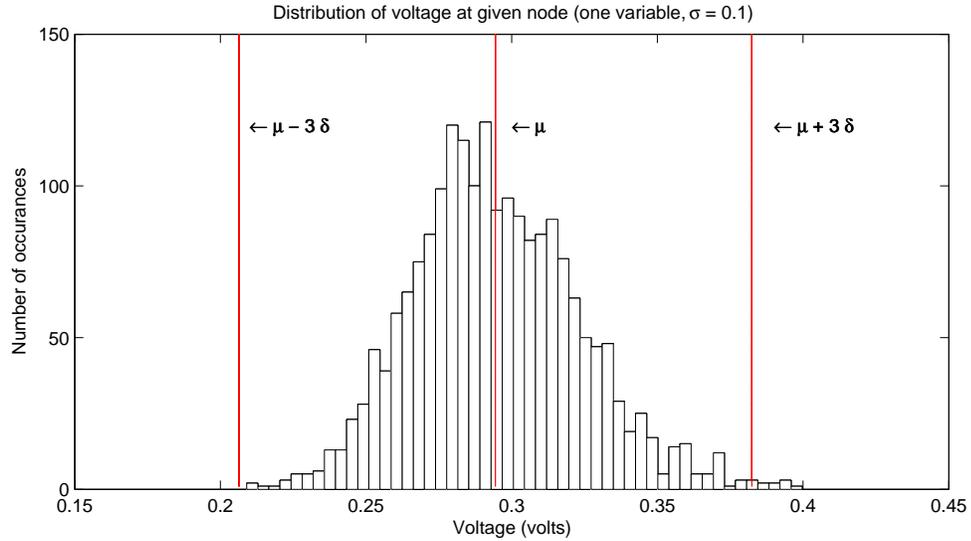
Figure 4.1: Distribution of the voltage in a given node with one Gaussian variable, $\sigma_g = 0.1$, at time 50ns when the total simulation time is 200ns.

Another observation is that when standard deviation, σ_g, is small, the shape
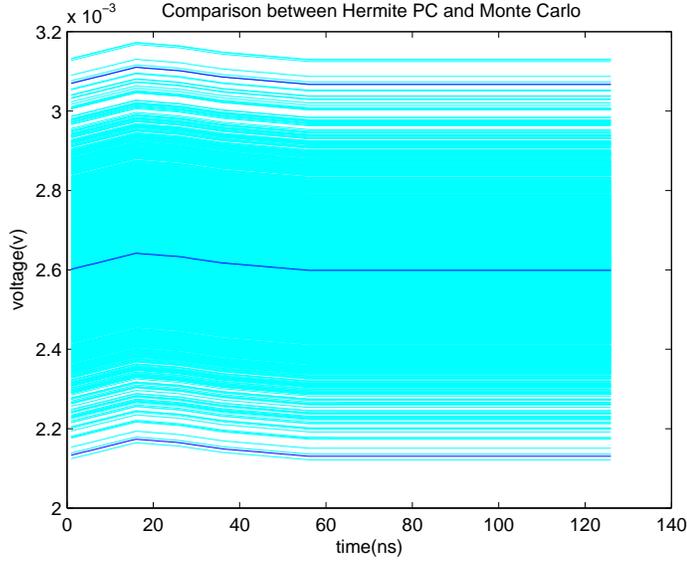
Figure 4.2: Distribution of the voltage caused by the leakage currents in a given node with one Gaussian variable, $\sigma_g = 0.5$, in the time instant from 0ns to 126ns.

looks like Gaussian as in Fig. 4.1, but it is log-normal indeed. In the case of two random variables with one large and the other small standard deviations, the larger one dominates, which shows the shape of log-normal as in Fig. 4.3.

To consider multiple random variables, we divide the circuit into several partitions. We first divide the circuit into two parts. Fig. 4.3 shows the node voltage of one node of a particular time instance of a ground network with 336 nodes with two independent variables. The standard deviations for two Gaussian variations are $\sigma_{g1} = 0.5$, $\sigma_{g2} = 0.1$. The 3σ variations are also marked in the figure.

Table 4.2 and table 4.3 shows the speedup of the Hermite PC method over Monte Carlo method with 2000 samples considering one and two random variables, respectively.

In two tables, $\#node$ is the number of nodes in the power grid circuits. p is

44
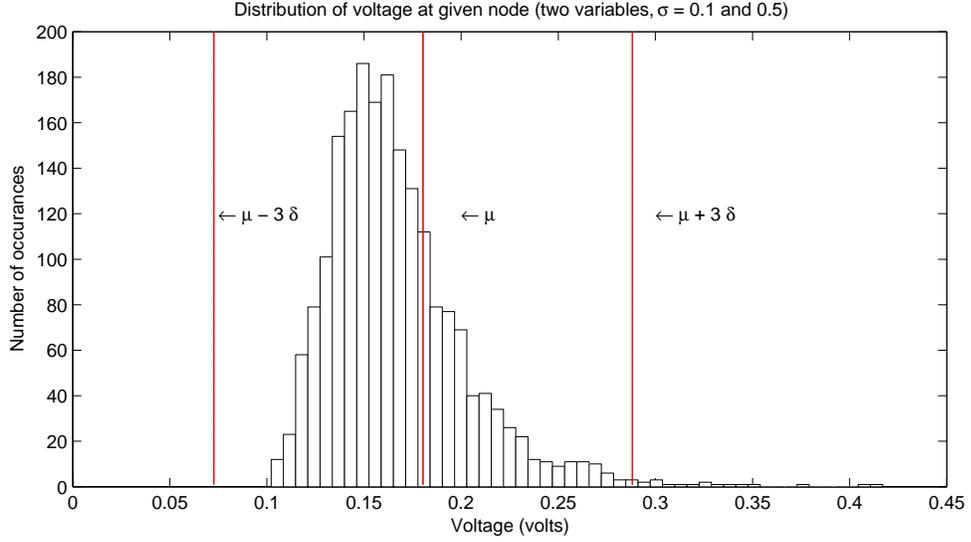
Figure 4.3: Distribution of the voltage in a given node with two Gaussian variables, $\sigma_{g1} = 0.1$ and $\sigma_{g2} = 0.5$, at time 50ns when the total simulation time is 200ns.

Ckt	#node	p	n	MC(s)	#MC	HPC(s)	Speedup
gridrc_6	280	2	1	766.06	2000	1.0156	754.3
gridrc_12	3240	2	1	4389	2000	8.3281	527.0
gridrc_5	49600	2	1	2.3×10^5	2000	298.02	771.76

Table 4.2: CPU time comparison with the Monte Carlo method of one random variable.

Ckt	#node	p	n	MC (s)	#MC	HPC (s)	Speedup
gridrc_3	280	2	2	1.05×10^3	2000	2.063	507.6
gridrc_5	49600	2	2	2.49×10^5	2000	445.6	558.7
gridrc_9	105996	2	2	6.11×10^5	2000	1141.8	535.1

Table 4.3: CPU time comparison with the Monte Carlo method of two random variables

the order of the Hermite PCs and n is the number of independent Gaussian random variables. $\#MC$ is the number of samples used for Monte Carlo method. HPC and MC represent the CPU times used for Hermite PC method and MC method respectively. It can be seen that the proposed method is about two order of magnitude faster than the MC method.

When more Gaussian variables are used for modeling intra-die variations, we need more Hermite PC coefficients to compute. Hence, the speedup will be smaller if the MC method uses the same number of samples as shown in *gridrc_12*. Also, one observation is that the speedup depends on the sampling size in MC method. The speedup of the proposed method over the MC method depends on many factors such as the order of polynomials, number of variables, etc. In general, speedup should not have a clear relationship with the circuit sizes. We still use 2000 samples for MC, which represent about 97.7% accuracy (as the error in MC is roughly $1/\sqrt{2000}$ for 2000 samples).

4.4.3   Examples with spatial correlation

To model the intra-die variations with spatial correlations, we divide the power grid circuit into several parts. We first consider that circuit is partitioned into two parts. In this case, we have two independent random current variables , ξ_1 and ξ_2. The correlated variables for the two parts are $\Phi_1 = \xi_1 + 0.5\xi_2$ and $\Phi_2 = \xi_2 + 0.5\xi_1$ respectively as shown in Fig. 4.4.

$\Phi_1 = \xi_1 + 0.5\xi_2$	$\Phi_2 = \xi_2 + 0.5\xi_1$

Figure 4.4: Correlated random variables setup in ground circuit divided into two parts.

ckt	#nodes	Mean		Std Dev	
		Non-PCA % error	PCA % error	Non-PCA % error	PCA % error
1	336	10.3	0.52	18.8	1.13
2	645	8.27	0.59	11.4	1.16
3	1160	10.8	0.50	2.6	0.73

Table 4.4: Comparison between non-PCA and PCA against Monte Carlo methods.

Table 4.4 shows the error percentage of mean and standard deviation of the comparison between *Monte Carlo* and *HPC* with Principal Component Analysis (PCA) and the comparison between *Monte Carlo* and *HPC* without PCA. As shown in the table, it is necessary to use PCA when spatial correlation is considered. Fig.4.5 shows the node voltage distribution of one certain node in a ground network with 336 nodes, using both PCA and non-PCA method.

Figure 4.5: Distribution of the voltage in a given node with two Gaussian variables with spatial correlation, at time 70ns when the total simulation time is 200ns.

To get more accuracy, we divide the circuit into four parts and each part has correlation with its neighbor as shown in Fig.4.6. ϕ is the correlated random variable

$\phi 1 = \zeta 1 + 0.5\zeta 2 + 0.5\zeta 3$	$\phi 3 = \zeta 3 + 0.5\zeta 1 + 0.5\zeta 4$
$\phi 2 = \zeta 2 + 0.5\zeta 1 + 0.5\zeta 4$	$\phi 4 = \zeta 4 + 0.5\zeta 2 + 0.5\zeta 3$

Figure 4.6: Correlated random variables setup in ground circuit divided into four parts.

vector we use in the circuit. $\zeta = [\zeta_1, \zeta_2, \zeta_3, \zeta_4]$ are independent Gaussian distribution random variables with standard deviations $\zeta_1 = 0.1$, $\zeta_2 = 0.2$, $\zeta_3 = 0.1$ and $\zeta_4 = 0.5$. Fig.4.7 is the voltage distribution of a given node. The mean voltage and voltages of worst case are given as the dashed line. Fig.4.8 is the voltage distribution of a circuit with 1160 nodes. The circuit is partitioned into 25 parts of five rows and five columns with spatial correlation. The dashed blue lines are mean, upper bound and lower bound by Hermite PC. While the solid red lines are mean, upper bound and lower bound by Monte Carlo of 2000 times.

Note that the size of the ground networks we analyzed is mainly limited by the solving capacity of Matlab on a single Intel CPU Linux workstation. Given long simulation time of large Monte Carlo sampling runs, we limit the ground network size to about 3000 nodes.

Also note that for more accurate modeling, we need to have more partitions of the circuits and thus more independent Gaussian variables are needed as shown in [10].

4.4.4 Consideration of variations in both wire and currents

Considering variation in conductance, capacitor and leakage current, Fig. 4.9 shows the node voltage distribution at one node of ground circuit, Circuit4, which contains 280 nodes. The maximum 3δ variation are 10% in ξ_g, ξ_c and ξ_I. In the figures, the
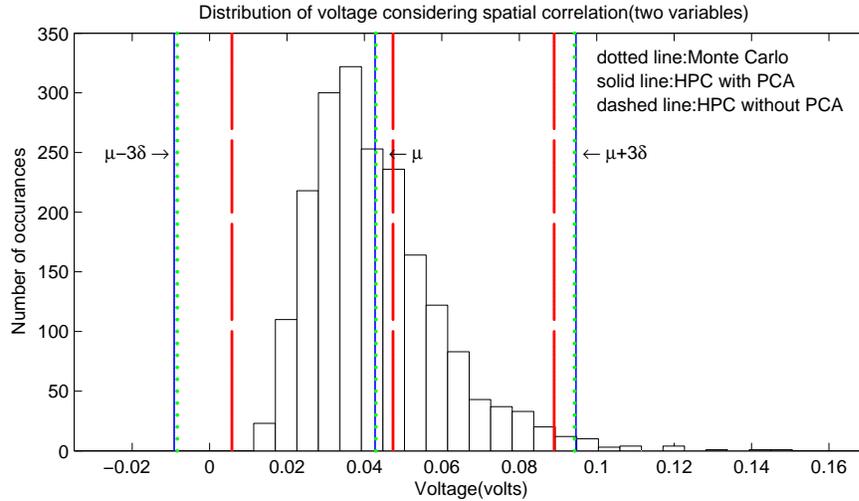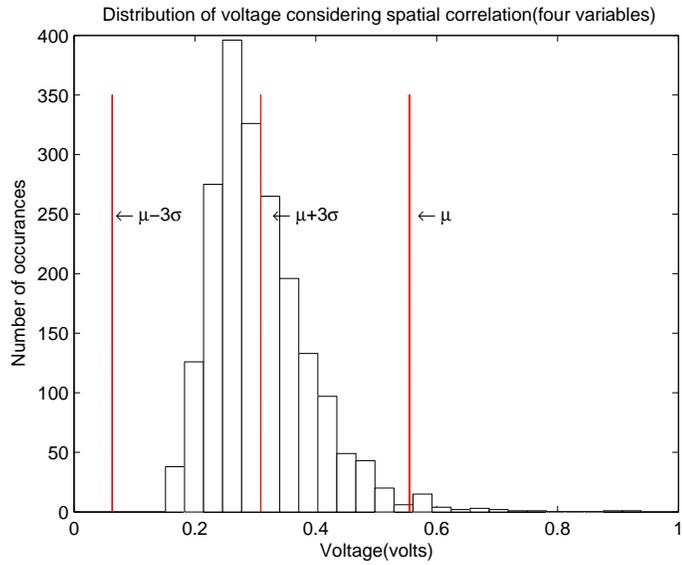
Figure 4.7: Distribution of the voltage in a given node with four Gaussian variables with spatial correlation, at time 30ns when the total simulation time is 200ns.
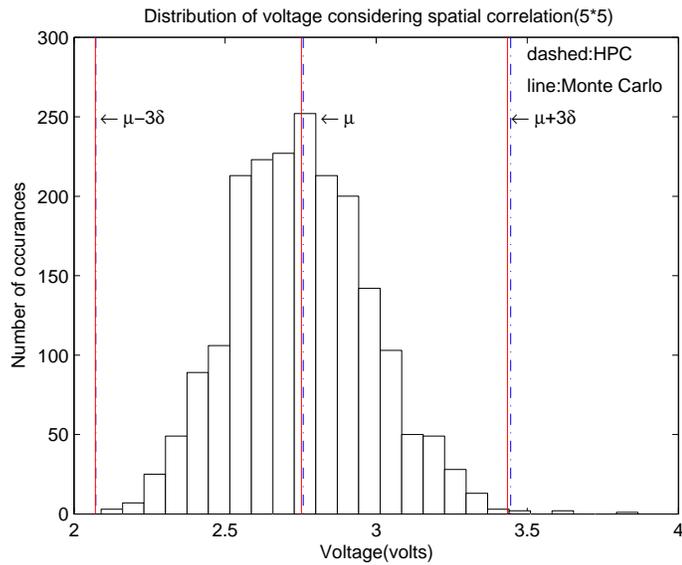
Figure 4.8: Distribution of the voltage in a given node with circuit partitioned of 5*5 with spatial correlation, at time 30ns when the total simulation time is 200ns.

solid lines are the mean voltage and worst case voltages using HPC method. The histogram bars are the Monte Carlo results of 2000 samples. The dotted lines are the mean voltage and worst case voltage of the 2000 samples. From the figures we can see that results getting from two methods match very well.
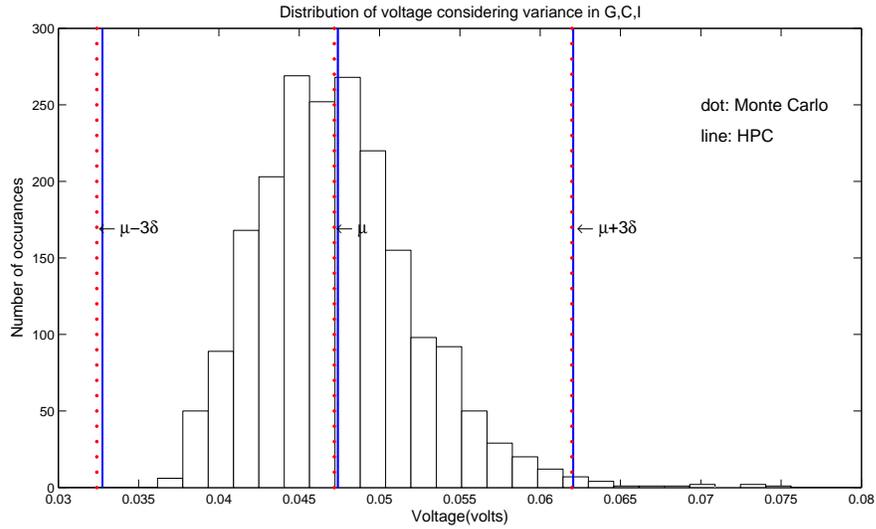
Figure 4.9: Distribution of the voltage in a given node in circuit5 with variation on G,C,I, at time 50ns when the total simulation time is 200ns.

Table 4.5 shows the CPU speedup of *HPC* method over *MC* method. The sample numbers of Monte Carlo is 3500 and we can see that the proposed method is about two orders of magnitudes faster than the Monte Carlo method when considering variations in conductance, capacitors and voltage sources. The speedup becomes smaller for larger circuits. This is because the super-linear time complexity of linear solver as the augmented matrices in Eq.4.45 grow faster than each individual matrices G_i and C_i. The proposed method does not favor very large circuits. Practically, this scalability problem can be mitigated by using partitioning-based strategies [13].

Ckt	#node	MC(s)	HPC(s)	Speedup
gridrc_6	280	1320.1	9.25	142.7
gridrc_12	3240	12183	141.4	86.2
gridrc_62	9964	63832	3261	19.6

Table 4.5: CPU time comparison with the Monte Carlo method considering variation in G,C,I.

4.5   Summary

In this section, we have proposed a new statistical simulation method for fast estimating the voltage variations from the process-induced log-normal leakage current variations with spatial correlations. The new analysis is based on the Hermite polynomial chaos (PC) representation of random processes. We extended the existing Hermite PC based power grid analysis method [23] by considering log-normal leakage distributions as well as the consideration of the spatial correlations. The new method considers both log-normal leakage distribution and wire variations at the same time. Our experimental results show that the new method is more accurate than the Gaussian-only Hermite PC using the Taylor expansion method for analyzing leakage current variations and two orders of magnitude faster than Monte Carlo methods with small variation errors. In the presence of spatial correlations, method without considering the spatial correlations may lead to large errors, roughly 8%-10% in our tested cases, if correlation is not considered. Experimental results show the correctness and high accuracy of the proposed method. It leads to about 1% or less of errors in both mean and standard deviations and is about two orders of magnitude faster than Monte Carlo methods. However, with the increase of the number of random variables, the size of the augmented matrices grows bigger very fast. It would be timing consuming then. Chapter5 will introduce an improved method.

Chapter 5

Fast Variational Simulation of Power Grids considering Process Variation

5.1 Problem Statement

In this chapter, we propose a new stochastic method for analyzing the voltage drop variations of on-chip power grid networks with log-normal leakage current variations. The new method, called *StoEKS*, still applies the spectral stochastic method to solve for the variational responses. But different from the existing spectral-stochastic-based simulation method, the extended Krylov subspace method (EKS) is employed to compute variational responses using the augmented matrices consisting of the coefficients of Hermite polynomials. Our work is inspired by recent spectral-stochastic-based model order reduction method [66]. We apply this work to the variational analysis

of on-chip power grid networks considering the variational leakage currents with the log-normal distribution.

Our contribution lies in the acceleration of the spectral stochastic method using the extended Krylov subspace method to fast solve the variational circuit equations for the first time. By using the Krylov-subspace based reduction technique, the new method partially mitigates the increased circuit-size problem associated with the augmented matrices from the Galerkin-based spectral stochastic method. We will show how the coefficients of Hermite PCs are computed for variational circuit matrices and for the current moments used in EKS with log-normal distribution. Experimental results show that the proposed StoEKS is about two order magnitude faster than the existing Hermite PC based simulation method, having similar error compared with Monte Carlo method. StoEKS can analyze much larger circuits than the existing Hermite PC method in the same computation platform.

5.1.1   Considering both wire and leakage variation

In this chapter, we assume that we have a number of independent (uncorrelated) transformed ortho-normal Gaussian random variables $\xi_i(\theta), i = 1, ..., n$, which actually model the channel length and the device threshold voltage variations. The spatial correlation in the intra-die variation can be processed by using the Karhunen-Loeve (K-L) transformation, principal component analysis method, which is the discretized version of K-L method [35], to transform the correlated variables into un-correlated variables before spectral statistical analysis [23, 38].

Let Θ denote the process sampling space. Let $\theta \in \Theta$, $\xi_i : \theta \to R$ denote a normalized Gaussian variable and $\xi(\theta) = [\xi_{1d}(\theta), ..., \xi_{rd}(\theta)]$ is a vector of r Gaussian variables.

After orthogonal transformation operation, we obtain independent random variable vectors $\xi = [\xi_1, ..., \xi_n]$. Notice that $n \leq r$ in general. The PCA method can be either used to strongly correlated dependent random variables or weak correlated random variables. The difference between the two circumstance is the number of independent variables, which is denoted as n here. After doing PCA, strongly correlated random variables, the independent random variable vector $\xi = [\xi_1, ..., \xi_n]$ are in small size. In other words, the number of random variables are reduced a lot. While for weak correlated, the number of independent random variables is not reduced a lot. Our method can deal with both strongly and weak correlated random variables.

Therefore, given the process variations, the MNA equation for Eq.2.1 becomes

$$G(\xi)v(t, \xi) + C(\xi)\frac{dv(t, \xi)}{dt} = Bu(t, \xi) \tag{5.1}$$

In this part, we assume that the varitional current source in Eq.5.1, $u(t, \xi)$, consists of two components:

$$u(t, \xi) = u_d(t) + u_v(t, \xi) \tag{5.2}$$

where $u_d(t)$ is the dynamic current vector from circuit switching, which are still modeled as deterministic currents as we only consider the leakage variations. $u_v(\xi, t)$ is the variational leakage current vector, which is dominated by sub-threshold leakage currents and it may change over time also. $u_v(t, \xi)$ follows the log-normal distribution.

The problem we need to solve is to efficiently find the mean and variance of voltage $u(t, \xi)$ at any node at any time instance without using the time-consuming sampling-based method, such as Monte-Carlo.

5.2 EKS review

5.2.1 Krylov subspace

In this subsection, we briefly introduce model order reduction used in Power Grid network simulation and the definition of Krylov subspace, from which comes the Extended Krylov Subspace(EKS).

One of the critical issue of power grid network is the large size of the network, sometimes it is over millions. It is slow to do the simulation and optimization of such power grid network. One of the solutions to this problem is to do model order reduction of the network with equivalent admittance and currents at the interface points as in Figure5.1. This approach works well since the reduction of the power grid network involves solving a symmetric positive-definite system.
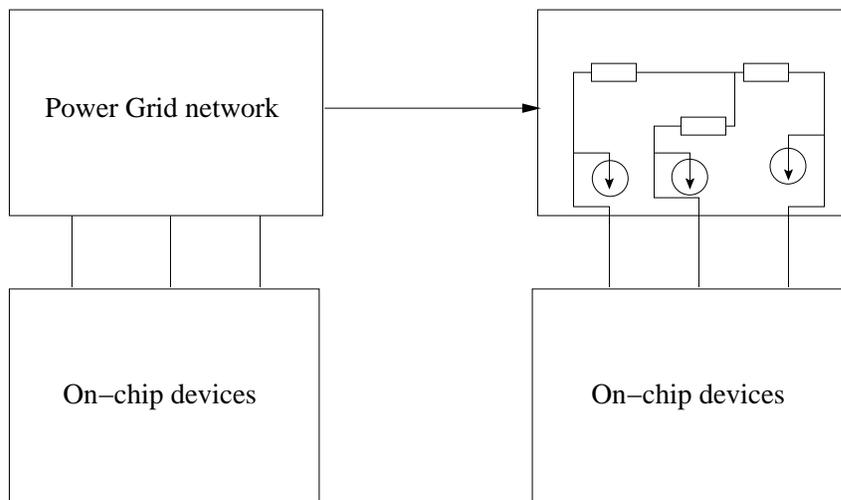
Figure 5.1: Moder Order Reduction used in Power Grid network simulation

One of the model order reduction method uses Krylov subspace as the basic to

reduce the system matrix. The definition of Krylov subspace is as follows:

$$\mathbf{K_r}(\mathbf{A}, \mathbf{b}) = \text{span}\{\mathbf{b}, \mathbf{Ab}, \mathbf{A^2b}, ..., \mathbf{A^{r-1}b}\} \tag{5.3}$$

$\mathbf{A}$ here is an n-by-n matrix and b is with dimension of n. Span here represents linear subspace spanned by the images of b under the first r powers of A.

Model order reduction method [43, 47] is to reduce the MNA matrix

$$C\dot{x}_n = -Gx_n + Bu_n$$

$$i_r = B^T x_n \tag{5.4}$$

Here, $C \in R^{n \times n}$, $G \in R^{n \times n}$ and $B \in R^{n \times r}$. We can use model order reduction method such as balanced truncated method [46] or Krylov-based model order reduction method [43] to reduce the dimension of C, G and B to $\tilde{C} \in R^{q \times q}$, $\tilde{G} \in R^{q \times q}$, $\tilde{B} \in R^{q \times r}$, while preserving the system characteristic such as moments matching of the system transfer function. Here q is much smaller than n.

The Krylov-based model order reduction method, PRIMA [43], for example, would use Krylov subspace to reduce the size of the matrix dumped from the circuit via MNA. In the MNA equation for power grid circuit as in Eq. 5.4, the Krylov subspace is written as

$$Kr(\mathbf{A}, \mathbf{R}, \mathbf{q}) \equiv \mathbf{colsp}[\mathbf{R}, \mathbf{AR}, \mathbf{A^2R}, ..., \mathbf{A^{k-1}R}, \mathbf{A^k r_0}, \mathbf{A^k r_1}, ..., \mathbf{A^k r_l}],$$

$$k = \lfloor q/N \rfloor, l = q - kN \tag{5.5}$$

Here, $A = -G^{-1}C$ and $R = G^{-1}B$. Based on this, $X \in R^{n \times q}$ is obtained, which is

an orthonormal matrix spanning the Krylov space.

Then, the system matrix A is reduced to a small block upper Hessenberg matrix H_q. The procedure involves successively filling in the columns of X in the relation $AX = XH_q$ subject to $X^T X = I_q$. Then, x_n can be obtained by

$$x_n = Xz_q \tag{5.6}$$

$z_q \in R^{q \times 1}$ here is the reduced-order system variable as

$$\tilde{C}\dot{z}_q = -\tilde{G}z_q + \tilde{B}u_n$$
$$i_r = \tilde{B}^T z_q \tag{5.7}$$

Here, $\tilde{C} \in R^{q \times q}$, $\tilde{G} \in R^{q \times q}$, $\tilde{B} \in R^{q \times r}$ obtained from

$$\tilde{C} = X^T C X, \tilde{G} = X^T G X, \tilde{B} = X^T B \tag{5.8}$$

5.2.2   Extended Krylov subspace

In this subsection, we briefly review the extended Krylov subspace (EKS) method in [61] and [32] for fast computation of responses from linear dynamic systems.

The EKS method uses the Krylov-like reduction method to speedup the simulation process. Different from the Krylov-based model order reduction method, EKS performs the reduction considering both system matrices and input signals before the simulation (so the subspace is no longer Krylov subspace). So it essentially is a simulation approach using the Krylov subspace reduction method. It assumes input

signals can be represented by piece-wise linear (PWL) sources.

Let $V = [\hat{v}_1, \hat{v}_2, ...\hat{v}_k]$ be an orthogonal basis for moment subspace $(m_0, m_1, ..., m_k)$ of input $u(t)$. Fig5.2 is the high-level description of the EKS algorithm [61].

Algorithm: EKS ALGORITHM
Input: G,C,B,$u(t)$ and moment order q
Output: Orthogonal basis $V =$ $\{\hat{v}_0, \hat{v}_2, ...\hat{v}_{q-1}\}$
1   $\hat{v}_0 = \alpha_0 v_0$, where $v_0 = G^{-1}Bu_0$, $\alpha_0 = \frac{1}{norm(v_0)}$;   2   set $h_s = 0$;   3   for $i = 1 : q - 1$ 4 $v_i = G^{-1}\{\Pi_{j=0}^{i-1}\alpha_j Bu_i - C(\hat{v}_{i-1} + \alpha_{i-1}h_s)\}$;   5 $h_s = 0$;   6   for $j = 0 : i - 1$ 7 $h = \hat{v}_j^T v_i$;   8 $h_s = h_s + h\hat{v}_j$;   9   end   10 $\bar{v}_i = v_i - h_s$;   11 $\alpha_i = \frac{1}{norm(\bar{v}_i)}$;   12 $\hat{v}_i = \alpha_i \bar{v}_i$ 13   end

Figure 5.2: The EKS algorithm.

Then the original circuit described by Eq.2.1 can be reduced to a smaller system

$$\tilde{G}z + \tilde{C}\frac{dz(t)}{dt} = \tilde{B}u \tag{5.9}$$

where

$$\tilde{G} = V^T G V,$$

$$\tilde{C} = V^T C V,$$

$$\tilde{B} = V^T B,$$

$$v(t) = V z(t)$$

After the reduced system in Eq.5.9 has been solved for the given input $u(t)$, the solution $z(t)$ can then be mapped back into original space by $v(t) = V z(t)$.

As the EKS models a PWL source as a sum of delayed ramps in Laplace domain, the terms, however, contain $1/s$ and $1/s^2$ moments [61], while the traditional Krylov space starts from $0th$ moment. Therefore, moment shifting must be made in EKS, which would cause complex computation and more errors. This problem is resolved in [32] in the IKES algorithm, which shows that the moments of $1/s$ and $1/s^2$ are zeros for PWL input sources.

Assume that we want to obtain a single input source $u_j(s)$ in the following moment form:

$$u_j(s) = u_1 + u_2 s + u_3 s^2 + ... + u_L s^{L-1}$$

A PWL source $u_j(t)$ is represented by a series of value-time pairs such as (a_1, τ_1), (a_2, τ_2),..., (a_{K+2}, τ_{K+2}) and L moments needed to be calculated. As proposed in [32], the mth moment for current source $u_j(t)$ in a current source vector $u(s)$ can be

calculated as

$$
\begin{aligned}
u_{j,m} &= (a_1 - \alpha_1 \frac{\tau_1}{m+1})\beta_1^{(m)} - \sum_{i=1}^{k}(\alpha_i - \alpha_{i+1})\beta_{i+1}^{(m+1)} \\
&\quad -(a_{K+2} - \alpha_{K+1}\frac{\tau_{k+2}}{m+1})\beta_{K+2}^{(m)}, \ m = 1, ..., L.
\end{aligned}
\tag{5.10}
$$

Here,

$$
\beta_i^{(m)} = \frac{(-\tau_i)^m}{m!}, \ \alpha_i = \frac{a_{i+1} - a_i}{\tau_{i+1} - \tau_i}
$$

The EKS/IEKS method, however, has its limitations. One major drawback is that current sources have to be represented in the explicit moment form, which may not be accurate and not numerical stable when high order moments are employed for high-frequency rich current waveforms owning to the well-known problem in the explicit moment matching method [48].

5.3   New Stochastic EKS method − StoEKS

In this section, we present the new stochastic simulation algorithm, StoEKS, which is based on both the spectral stochastic method and the extended Krylov subspace method. The main idea is that we use the spectral stochastic method to convert the statistical simulation into a deterministic simulation problem. Then we apply EKS to solve the converted problem.

5.3.1 StoEKE algorithm flowchart

First, we present StoEKS algorithm flowchart, which is shown in Fig. 5.3. The algo-
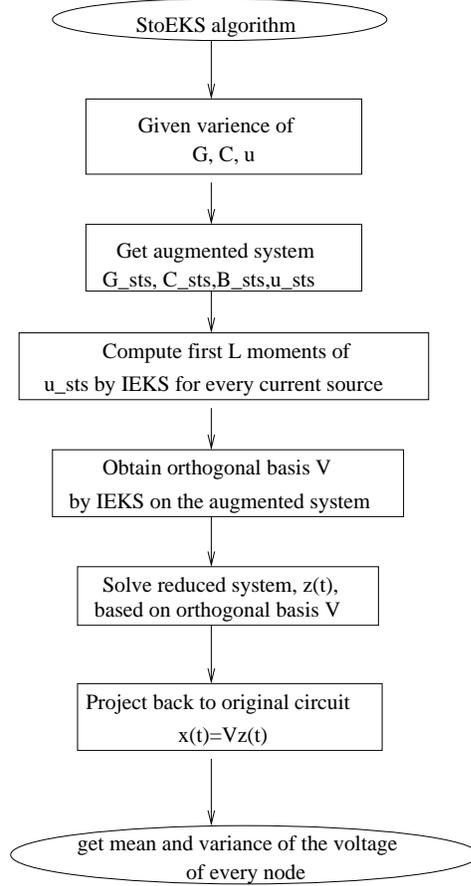
Figure 5.3: Flowchart of the StoEKS algorithm

rithm starts with variational $G(\xi)$, $C(\xi)$ and variational input source $u(t, \xi)$. Then, it applies spectral stochastic method to convert the variational system Eq.2.2 into a deterministic system, which consists of augmented matrices of $G(\xi)$ and $C(\xi)$ and position matrix B in Eq.2.2 with new unknowns. Then we generate the first L moments of coefficients of Hermite polynomial of current sources, U_L with log-normal distribution. Finally, we apply EKS/IEKS to solve the obtained deterministic system

for response Z using the computed projection matrix V. After this we get back to the transient response of the original augmented system by $v(t) = Vz(t)$. Finally we compute the mean and variance of any voltage nodes from $v(t)$.

In the following subsections, we present the detailed descriptions for some critical steps of the StoEKS algorithm.

5.3.2 Generation of the augmented circuit matrices

We first show how we convert the variational circuit equation into a deterministic one, which is suitable for EKS. Our work follows the recent proposed stochastic model order reduction (SMOR) method [66]. SMOR is based on Hermite Polynomial Chaos and the Krylov-based projection method.

We first assume that $G(\xi)$, $C(\xi)$ and $u(t, \xi)$ in Eq.5.1 are represented in Hermite PC forms with a proper order P:

$$G(\xi) = G_0 + G_1 H_1(\xi) + G_2 H_2(\xi) + \dots + G_{P-1} H_{P-1}(\xi)$$

$$C(\xi) = C_0 + C_1 H_1(\xi) + C_2 H_2(\xi) + \dots + C_{P-1} H_{P-1}(\xi)$$

$$u(t, \xi) = (u_0(t) + u_d(t)) + u_1(t) H_1(\xi) + \dots + u_{P-1}(t) H_{P-1}(\xi)$$

Here, $H_i(\xi)$ are the Hermite PC basis functions for $G(\xi)$, $C(\xi)$ and $u(t, \xi)$. P is also the number of these basis functions, which depends on the number of random variables n and the expansion order p in Eq.3.7. G_i, C_i and u_i are the Hermite polynomial coefficients of conductance, capacitors and current source. G_0 and C_0 are the mean value of conductance and capacitors. G_i and C_i are variational part for conductance and capacitors.

Ideally, to obtain the G and C in the HPC format, i.e., to compute G_i and C_i from the width, length variables, one can use stochastic spectral analysis method [30], which is a fast Monte-Carlo method or other extraction methods. For this paper, we simply assume that we obtain such information, The detail of how G_i and C_i are obtained as follows:

$$G_i = a_i * G_0, C_i = a_i * C_0, i = 1, ..., P - 1 \tag{5.11}$$

a_i is the variational percentage for H_i.

Substitute Eq.5.11 into Eq.5.1, the system equations become:

$$\sum_{i=0}^{P-1}\sum_{j=0}^{P-1} G_i v_j H_i H_j + s \sum_{i=0}^{P-1}\sum_{j=0}^{P-1} C_i v_j H_i H_j =$$
$$u_d(t) + \sum_{i=0}^{P-1} u_i(t) H_i \tag{5.12}$$

Here, v_i are the coefficients of Hermite Polynomial of node voltages $v(t, \xi)$ as:

$$v(t, \xi) = v_0(t) + v_1(t) H_1 + v_2(t) H_2 + ... + v_{P-1}(t) H_{P-1} \tag{5.13}$$

After performing the inner product of H_k on both sides of the equation Eq.5.12, it will become

$$\sum_{i=0}^{P-1}\sum_{j=0}^{P-1} G_i v_j < H_i H_j, H_k > + s \sum_{i=0}^{P-1}\sum_{j=0}^{P-1} C_i v_j < H_i H_j, H_k >$$
$$= \sum_{i=0}^{P-1} u_i < H_i, H_k > + < H_k, 1 > v_d(t), \quad k = 0, 1, ..., P - 1 \tag{5.14}$$

where $< H_i H_j, H_k >$ is the inner product of $H_i H_j$ and H_k. On the right hand side of Eq.5.14, the inner product is calculated based on H_i and H_k.

Notice that $< H_k, 1 >= 1$, when $k = 0$; $< H_k, 1 >= 0$ when $k \neq 0$. In general, the coefficients of $H_i H_j$ are calculated in Eq.5.12 and the inner product is defined as

$$< H_i H_j, H_k >= \int_{-\infty}^{+\infty} H_i H_j H_k \mathrm{d}\xi \tag{5.15}$$

considering the independent of Hermite Polynomial H_i, H_j and H_k. Also, the inner product is similar for

$$< H_i, H_j >= \int_{-\infty}^{+\infty} H_i H_j \mathrm{d}\xi \tag{5.16}$$

The inner product is a constant and can be computed a priori and stored in a table for fast computation. Based on the P equations and the orthogonal nature of the Hermite polynomials, these equations can be written in matrix form as

$$(G_{sts} + sC_{sts})V = B_{sts}u_{sts} \tag{5.17}$$

$$G_{sts} = \begin{bmatrix} G_{00} & \dots & G_{0P-1} \\ \vdots & \ddots & \vdots \\ G_{P0} & \dots & G_{P-1P-1} \end{bmatrix},$$

$$C_{sts} = \begin{bmatrix} C_{00} & \dots & C_{0P-1} \\ \vdots & \ddots & \vdots \\ C_{P-10} & \dots & C_{P-1P-1} \end{bmatrix},$$

$$u_{sts} = \begin{bmatrix} u_0(t) + u_d(t) \\ u_1(t) \\ \vdots \\ u_{P-1}(t) \end{bmatrix}, V = \begin{bmatrix} V_0(t) \\ V_1(t) \\ \vdots \\ V_{P-1}(t) \end{bmatrix}, \tag{5.18}$$

$$B_{sts} = \begin{bmatrix} B_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & B_{P-1} \end{bmatrix} \tag{5.19}$$

$$B_i = B, G_{kj} = \sum_{i=0}^{P-1} G_i < H_i H_j, H_k >, C_{kj} = \sum_{i=0}^{P-1} C_i < H_i H_j, H_k >$$

where $G_{sts} \in R^{mP \times mP}$, $C_{sts} \in R^{mP \times mP}$ $B_{sts} \in R^{mP \times l}$, m is the size of the original circuit and P is the number of Hermite polynomials. In [66], PRIMA-like reduction is performed on Eq.5.17 to obtain the reduced variational system.

5.3.3 Computation of Hermite PCs of current moments with log-normal distribution

In this section, we show how to compute the Hermite coefficients for the variational leakage currents and their corresponding moments used in the augmented equation Eq.5.17.

Let $u_v^i(t, \xi)$ be the ith current in the current vector $u_v(t, \xi)$ in Eq.5.2, which is a function of the normalized Gaussian random variables $\xi = [\xi_1, \xi_2, ..., \xi_n]$ and time t.

$$u_v^i(t, \xi) \sim e^{g(t,\xi)} = e^{\sum_{j=0}^{n} g_j(t)\xi_j} \tag{5.20}$$

The leakage current sources are therefore following lognormal distribution. We can then present $u_v^i(t, \xi)$ by using Hermite PC expansion form:

$$
\begin{aligned}
u_v^i(t, \xi) &= \sum_{k=0}^{P} u_{vk}^i(t) H_k^n(\xi) \\
&= u_{v0}^i(t)\left(1 + \sum_{i=1}^{n} \xi_i g_i(t) + \right. \\
&\quad \left. \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{(\xi_i\xi_j - \delta_{ij})}{< (\xi_i\xi_j - \delta_{ij})^2 >} g_i(t)g_j(t) + ... \right)
\end{aligned}
\tag{5.21}
$$

where

$$u_{v0}^i(t) = e^{g_0(t) + \frac{1}{2}\sum_{i=1}^{n} g_i(t)^2}, P = \sum_{k=0}^{p} \frac{(n-1+k)!}{k!(n-1)!} \tag{5.22}$$

n is the number of random variables and p is the order of Hermite PC expansion.

As a result, the variational variable $u(t, \vec{\xi})$ leads to the u_{sts} in Eq.5.17:

$$u_{sts} = [u_0(t)^T + u_d(t)^T, u_2(t)^T, ..., u_{P-1}(t)^T]^T \tag{5.23}$$

Note that $u_d(t)$ is the deterministic current source vector.

In the EKS method, we need to compute the moments of input sources in frequency domain. Suppose (a_{i1}, τ_{i1}), (a_{i2}, τ_{i2}),..., (a_{iK+2}, τ_{iK+2}) are PWL series of value-time pairs for $u_i(t)$ or $u_0(t) + u_d(t)$ in Eq.5.23. Using equation Eq.5.10, we can get the first L moments for each u_i, $i = 1, 2, ..., P$ in Eq.5.23 respectively, and we have

$$u_i(s) = m_{u_{i1}} + m_{u_{i2}}s+, ..., m_{u_{iL}}s^{L-1} \tag{5.24}$$

where $m_{u_{ik}}$ is the kth order moment vector of Hermite PCs coefficient for u_i. In this way, we can compute the moments of Hermite PC coefficients for every current sources.

5.3.4 The StoEKS algorithm

Given the G_{sts}, C_{sts} and u_{sts} in moment forms, we can obtain the orthogonal V using the EKS algorithm. The reduced systems then can be obtained by these orthogonal basis V from equation Eq.5.10. The reduced system will become

$$\tilde{G}_{sts}z(t) + \tilde{C}_{sts}\frac{dz(t)}{dt} = \tilde{B}_{sts}u_{sts} \tag{5.25}$$

Here,

$$\tilde{G}_{sts} = V^T G_{sts} V, \tilde{C}_{sts} = V^T C_{sts} V, \tilde{B}_{sts} = V^T B_{sts} \tag{5.26}$$

The reduced system can be solved in the time domain by any standard integration algorithm. The solution of the reduced system, $z(t)$, can then be projected back to original space by $\tilde{v}(t) = V z(t)$.

By solving the augmented equation in Eq.5.17, we can obtain mean and variance of any node voltage $v(t)$ by

$$
\begin{aligned}
E(v(t)) &= E(v_0(t) + \sum_{i=1}^{P-1} v_i(t) H_i) = v_0 \\
var(v(t)) &= var(v_0(t) + \sum_{i=1}^{P-1} v_i(t) H_i) = \sum_{i=1}^{P-1} v_i(t)^2 var(H_i)
\end{aligned}
$$

Further the distribuation of $v(t)$ can also be easily calculated by the characteristic of Hermite PC and the distribution of $\xi_1, \xi_2, ..., \xi_N$. Fig. 5.4 is the StoEKS algorithm for given G_{sts}, C_{sts}, B_{sts}, u_{sts}.

5.3.5 A walk-through example

In the following, we consider a simple case where we only have three independent variables to illustrate the method. We assume that there are three independent variables ξ_g, ξ_c and ξ_I associated with matrices G and C and input sources respectively in the circuit.

We assume that the variational component in Eq.5.2. $u_v(t, \xi_I)$, follows lognormal

Algorithms: STOEKS
Input: augmented system G_{sts}, C_{sts}, B_{sts}, u_{sts}
Output: the HPC coefficients of node voltage, v
1 Get the first L moments of u_{sts} for each current source 2 Compute the orthogonal basis of subspace from Eq.5.17 V 3 Obtain the reduced system matrix from $\hat{G} = V^T G_{sts} V$, $\hat{C} = V^T C_{sts} V$, $\hat{B} = V^T B_{sts}$ 4 Solve $\hat{G}z(t) + \hat{C}\frac{dz(t)}{dt} = \hat{B}u_{sts}(t)$ 5 Project back to original space to get v(t) = Vz(t) 6 Compute the variational values (means, variance) of the specified nodes

Figure 5.4: The proposed StoEKS algorithm.

distribution as

$$u_v(t, \xi_I) = e^{g(t,\xi_I)}, g(t, \xi) = \mu_I(t) + \sigma_I(t)\xi_I \tag{5.27}$$

Then equation Eq.2.2 becomes

$$G(\xi_g)v(t, \xi) + C(\xi_c)\frac{dv(t, \xi)}{dt} = Bu(t, \xi_I) \tag{5.28}$$

$\xi = [\xi_g, \xi_c, \xi_I]$. The variation in width W and thickness T will cause variation in conductance matrix G and storage matrix C while variation in threshold voltage will cause variation in leakage currents $u(t, \xi_I)$. Thus, the resulting system can be written

as [23]:

$$G(\xi_g) = G_0 + G_1\xi_g, C(\xi_c) = C_0 + C_1\xi_c \tag{5.29}$$

G_0, C_0 represents the deterministic component of conductance and capacitance of the wires. G_1, C_1 represents sensitivity matrices of the conductance and capacitance. ξ_g, ξ_c are random variables with normalized Gaussian distribution, representing process variation in wires of conductance and capacitor, respectively.

ξ_I is a normalized Gaussian distribution random variable representing variation in threshold voltage.

Using Galerkin method as in [39] with second-order Hermite PCs, we end up solving the following equation

$$G_{sts}v(t) + C_{sts}\frac{dv(t)}{dt} = B_{sts}u_{sts}(t) \tag{5.30}$$

where

$$G_{sts} = \begin{bmatrix} G_0 & G_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ G_1 & G_0 & 0 & 0 & 2G_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & G_0 & 0 & 0 & 0 & 0 & G_1 & 0 & 0 \\ 0 & 0 & 0 & G_0 & 0 & 0 & 0 & 0 & G_1 & 0 \\ 0 & G_1 & 0 & 0 & G_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & G_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & G_0 & 0 & 0 & 0 \\ 0 & 0 & G_1 & 0 & 0 & 0 & 0 & G_0 & 0 & 0 \\ 0 & 0 & 0 & G_1 & 0 & 0 & 0 & 0 & G_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & G_0 \end{bmatrix}$$

$$(5.31)$$

$$
C_{sts} = \begin{bmatrix}
C_0 & 0 & C_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & C_0 & 0 & 0 & 0 & 0 & 0 & C_1 & 0 & 0 \\
C_1 & 0 & C_0 & 0 & 0 & 2C_1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & C_0 & 0 & 0 & 0 & 0 & 0 & C_1 \\
0 & 0 & 0 & 0 & C_0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & C_1 & 0 & 0 & C_0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & C_0 & 0 & 0 & 0 \\
0 & C_1 & 0 & 0 & 0 & 0 & 0 & C_0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C_0 & 0 \\
0 & 0 & 0 & C_1 & 0 & 0 & 0 & 0 & 0 & C_0
\end{bmatrix}
\tag{5.32}
$$

$$
u_{sts}(t) = [u_0(t) + u_d(t), 0, 0, u_3(t), 0, 0, u_6(t), 0, 0, 0]^T
$$

One observation we have is that although the augmented circuit matrices are much bigger than before, they are very sparse and also consist of repeated coefficient matrices from the HPC. As a result, the reduction techniques can significantly improve the simulation efficiency.

5.3.6 Computional complexity analysis

In this subsection, we analyze the computing costs for both StoEKS and HPC methods and show the theoretical advantage of StoEKS over the non-reduction based HPC method.

First, if the PCA operation is performed, which essentially uses singular value decomposition(SVD) on the covariance matrix, its computation cost is $O(ln^2)$. Here l is the number of original correlated random variables and n is the first n dominant singular values, which is also the number of independent random variables after PCA. Since the random viable l is typically much smaller than the circuit size, the running time of PCA is is not significant for the total cost.

After we transform the original circuit matrices into the augmented circuit matrices in Eq.5.17, which are still very sparse, the matrix sizes grow from $m \times m$ to $Pm \times Pm$, where P is the number of Hermite polynomials used. The number is dependent on the Hermite polynomial order and the number of variable used as shown in Eq.3.7.

Typically solving a $n \times n$ linear matrix takes $O(n^\alpha)$ (typically, $1 \le \alpha \le 1.2$ for sparse circuits), and matrix factorizations take $O(n^\beta)$ (typically, $1.1 \le \beta \le 1.5$ for sparse circuits). For HPC, assuming that we need to compute w time steps in transient analysis (taking w forward and backward substitutions after one LU decomposition), the computing cost then is

$$O(w(mP)^\alpha + (mP)^\beta). \tag{5.33}$$

While for StoEKS, we only need to approximately take q, the order of the reduced model, steps (after the one LU decomposition) to compute the projection matrix V. So the total computational cost is

$$O(q(mP)^\alpha + (mP)^\beta + mPq^2 + q^3 + wq^2). \tag{5.34}$$

without considering the cost of the PCA operations (ln^2) as we did not perform the PCA in our experiments. The last three items are the costs of performing the reductions (QR operation) and transient simulation of the reduced circuit (which have very dense matrices) in time domain. Since $q \ll w$, the computing cost of StoEKS can be significant lower than HPC. Also the proposed method can be further improved by using the hierarchical EKS method [9].

5.4 Experimental results

This section describes the simulation results of circuits with both capacitance, conductance variation and leakage current variation. The leakage current variation follows log-normal distribution. The capacitance and conductance variations follows Gaussian distribution.

All the proposed methods have been implemented in Matlab 7.0. All the experimental results are carried out on a Dell PowerEdge 1900 workstation (using a Linux system) with Intel Quadcore Xeon CPUs with 2.99Ghz and 16GB memory. To solve large circuits in Matlab, an external linear solver package UMFPACK [1] has been used, which is linked with Matlab using Matlab mexFunction.

We assume that the random variables used in the paper for G and C and current sources are independent after the PCA transformation.

First, we assume a time-variant leakage model, in which we assume that $u_v^i(t, \xi)$ in Eq.5.20 is a function of time t and further assume that $g_j(t)$, the standard deviation, is a fixed percentage, say 10%, of $v_d(t)$ in Eq.5.2, i.e. $g_i(t) = 0.1u_{di}(t)$, where $u_{di}(t)$ is the ith component of the PWL current $v_d(t)$.

Fig. 5.5, Fig. 5.6 and Fig. 5.7 show the results at one particular node under this configuration.

Figure 5.5: Distribution of the voltage variations in a given node by StoEKS, HPC and Monte Carlo of a circuit with 280 nodes with 3 random variables. $g_i(t) = 0.1u_{di}(t)$

Fig. 5.5 shows the node voltage distribution at one node of a ground network with 280 nodes, considering variation in conductance, capacitance and leakage current (with three random variables). The standard deviation (s.d.) of the log-normal current sources with one Gaussian variable is $0.1u_{di}(t)$. The s.d. in conductance and capacitance are also 0.1 of the mean. The mean and s.d. computed by the Hermite PC method, Hermite PC with EKS are also marked in the figure, which fit very well with the Monte Carlo (MC) results. In Fig. 5.5, the dotted lines are the mean and s.d. calculated by MC. The solid lines are the mean and s.d. by the algorithm [38], which is named as *HPC*. The dashed line are the results from the new StoEKS. The Monte Carlo results are obtained by 3000 samples. The reduced order for EKS is 5.
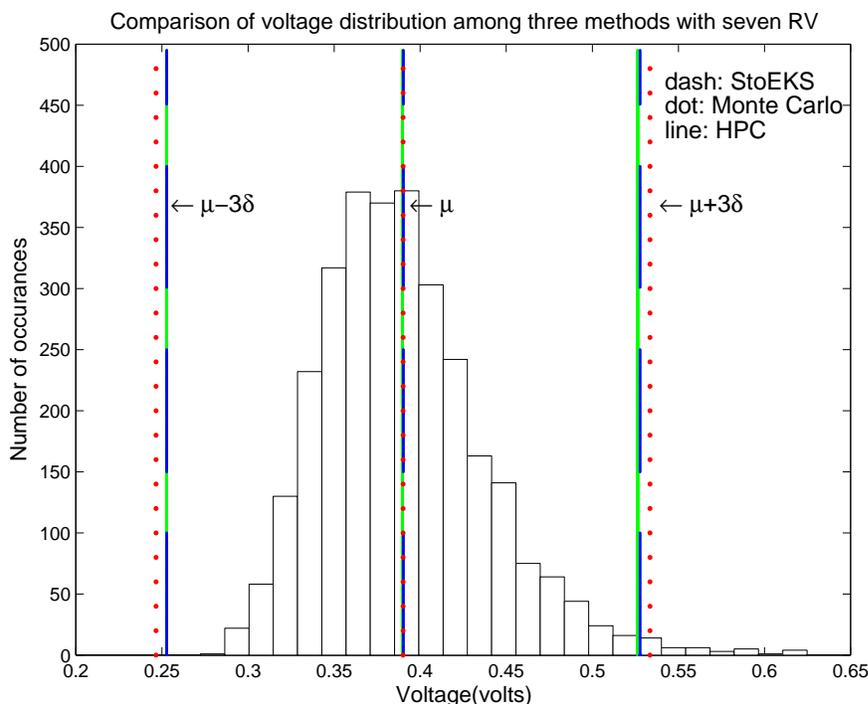
Figure 5.6: Distribution of the voltage variations in a given node by StoEKS, HPC and Monte Carlo of a circuit with 2640 nodes with 7 random variables. $g_i(t) = 0.1u_{di}(t)$

Fig. 5.6 shows the distribution at one node of a ground network with 2640 nodes. The parameter $g_i(t)$ value is set to the same as the ones in the circuit with 280 nodes. The s.d. in conductance are 0.02, 0.05 and 0.1 of the mean for three variables. The s.d. in capacitance are 0.02, 0.02 and 0.1 of the mean for three variable. There are totally seven random variables. The dotted lines represent the Monte Carlo results. And the dashed lines represent the results given by StoEKS. From these two figures, we can only see marginal difference between the three different methods. The reduced order for EKS is also 5, $q = 5$.

Fig. 5.7 shows the distribution at one node of a ground network with 280 nodes. But the variation setting of parameters is different. The standard divisions in con-
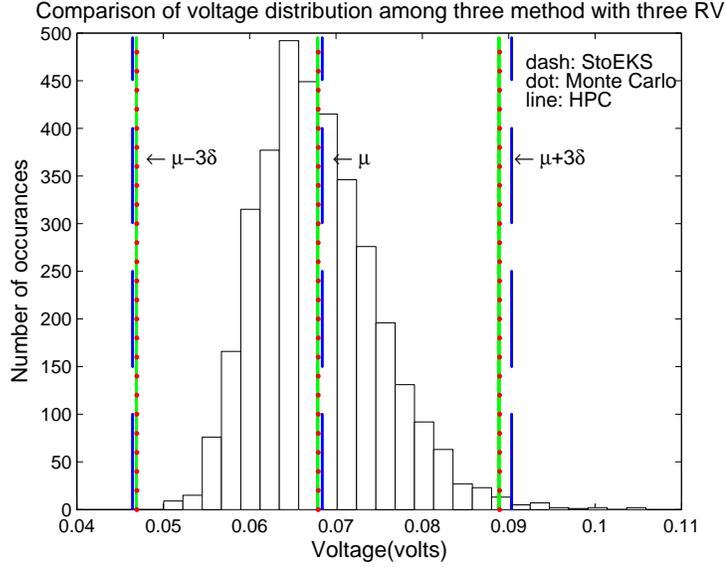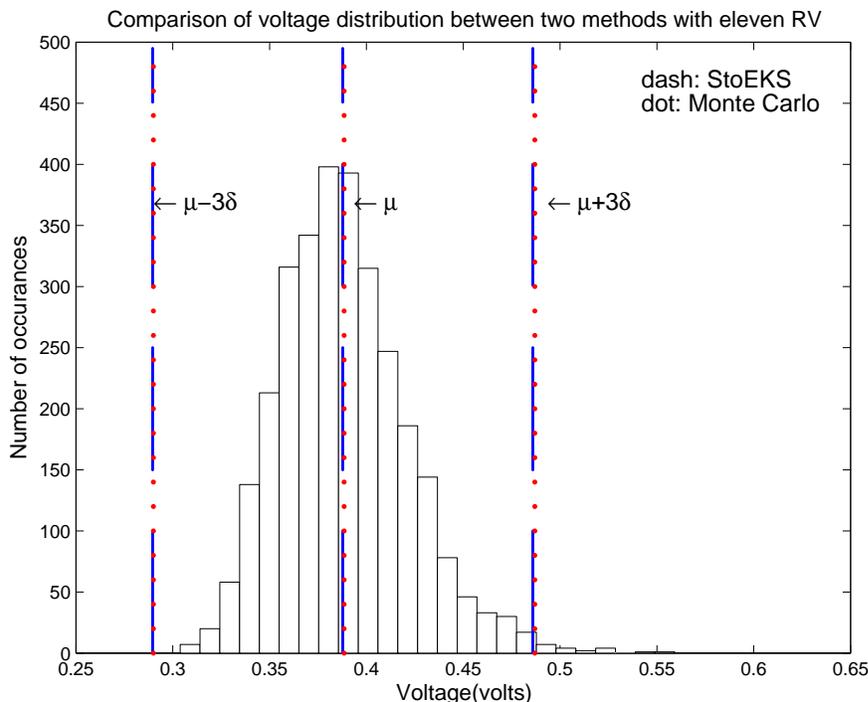
Figure 5.7: Distribution of the voltage variations in a given node by StoEKS and Monte Carlo of a circuit with 2640 nodes with 11 random variables. $g_i(t) = 0.1u_{di}(t)$

ductance are set to 0.02, 0.02, 0.03, 0.05 and 0.05 of the mean for five variables respectively, i.e their a_1 in Eq.5.11 are set to those values. The standard divisions in capacitances are set to 0.02, 0.03, 0.04, 0.05 and 0.05 of the mean for five variables respectively also. The standard deviation of the log-normal current sources is 0.1 of the mean. There are eleven random variables in all. It is even harder for HPC to compute mean and s.d. of the circuit. The dotted lines represent the Monte Carlo results. And the dashed lines represent the results given by StoEKS. The reduced order for EKS is 10.

Table 5.1 shows the speedup of the StoEKS and HPC methods over Monte Carlo method under different number of random variables. In the table, $\#RV$ is the number

#nodes	#RV	MC	StoEKS	speedup	HPC [38]	speedup
280	3	694.35	0.3	2314.5	2.37	292.97
280	7	671.46	2.37	283.31	227.94	2.94
280	11	684.88	24.26	28.23	914.34	0.74
2640	3	5925.7	4.33	1368.5	55.35	107.1
2640	7	5927.6	25.02	236.9	1952.2	3.04
2640	11	6042.2	693.27	8.72	−	−
12300	3	3.54×10^4	21.62	1637.4	298.84	118.5
12300	7	3.30×10^4	151.71	217.65	−	−
119600	3	−	258.21	−	−	−
119600	7	−	2074.8	−	−	−
1078800	3	−	1830.4	−	−	−

Table 5.1: CPU time comparison of StoEKS and HPC with the Monte Carlo method. $g_i(t) = 0.1u_{di}(t)$

of random variables used. In the table there are $3, 7, 11$ random variables. The variation value setup of 3 random variable is the same as the circuit used in Fig. 5.5. The variation value setup of 7 random variable is the same as the circuit used in Fig. 5.6. The variation value setup of 11 random variable is the same as the circuit used in Fig. 5.7. The first *speedup* is the speedup of StoEKS over MC and the second *speedup* is the speedup of StoEKS over HPC.

From the table, we observe that, we can't obtain the results from HPC or MC when the circuit becomes large enough in reasonable time. Meanwhile, StoEKS can deliver all the results. However, when the number of random variable becomes large, because the reduction process of the augmented system takes time, the speedup of StoEKS is not much as small number of random variables with MC.

We remark that the intra-die variations are typically very spatially correlated [12]. After the transformation like principal component analysis (PCA), the number of variables can be significantly reduced. As a result, in our examples, we do not assume

#nodes	#RV	Mean			Std Dev		
		MC	StoEKS	HPC	MC	StoEKS	HPC
280	3	0.047	0.047	0.047	0.0050	0.0048	0.0048
2640	3	0.39	0.39	0.39	0.048	0.046	0.046
12300	3	1.66	1.66	1.66	0.16	0.17	0.17
280	7	0.047	0.047	0.047	0.0056	0.0055	0.0055
2640	7	0.39	0.39	0.39	0.048	0.046	0.046
12300	7	2.56	2.56	−	0.31	0.30	−
280	11	0.047	0.047	0.047	0.0039	0.0039	0.0040
2640	11	0.39	0.39	−	0.033	0.033	−

Table 5.2: Accuracy comparison of different methods, StoEKS, HPC and Monte Carlo. $g_i(t) = 0.1u_{di}(t)$

large number of variables.

Table 5.2 and Table 5.3 show the mean and s.d. comparison of different methods over the MC method for several circuits. Again #RV is the number of random variables used. Table 5.2 contains the values we obtain from different methods and Table 5.3 presents the error comparison of *StoEKS* and *HPC* over Monte Carlo, respectively. We can see that StoEKS only has marginal difference from Monte Carlo while it is able to perform simulation on much larger circuit than the existing *HPC* method on the same platform.

Finally, we use a time-invariant leakage model, in which we assume that $u_v^i(\xi)$ in Eq.5.20 is not a function of time t and further assume that g_j, which is the standard deviation, is a fixed percentage, of a constant current value in Eq.5.2. In our test cases, we use the peak current, $I_p \approx 41mA$ as shown in Fig. 5.8, as the constant value. Fig. 5.9 shows the results in this configuration.

#nodes	#RV	StoEKS % Error in μ	HPC % Error in μ	StoEKS % Error in σ	HPC % Error in σ
280	3	0.19	0.28	3.14	3.10
2640	3	1.23	1.05	4.31	4.51
12300	3	0.10	0.08	2.95	2.98
280	7	0.063	0.17	1.12	1.54
2640	7	0.076	0.11	4.18	4.60
12300	7	0.23	–	0.23	–
280	11	0.42	0.21	0.18	0.52
2640	11	0.18	–	0.30	–

Table 5.3: Error comparison of StoEKS and HPC over Monte Carlo methods. $g_i(t) = 0.1u_{di}(t)$
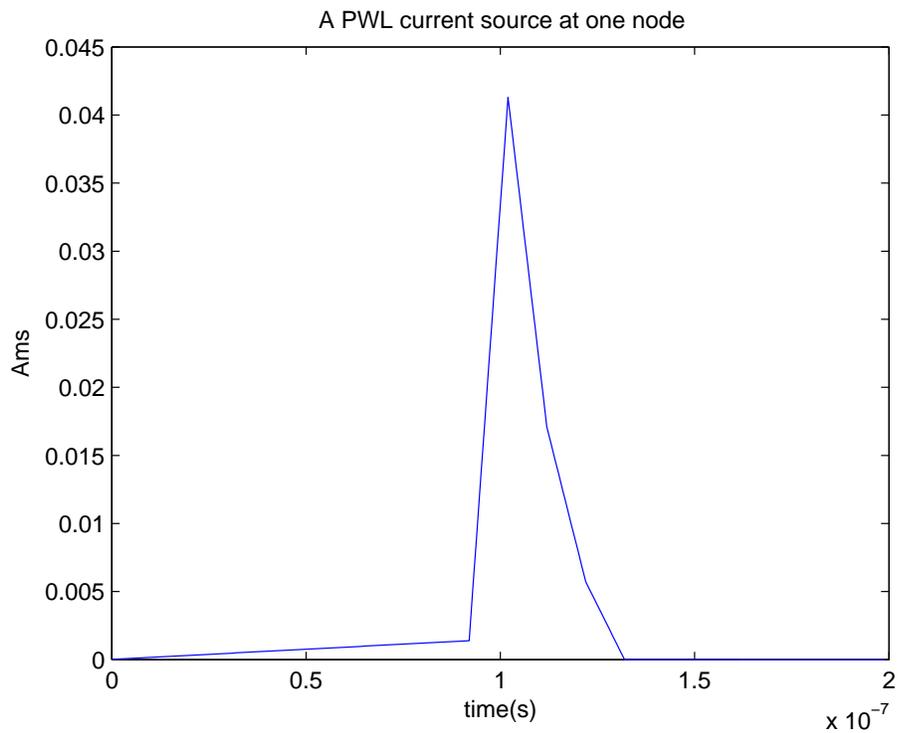
Figure 5.8: A PWL current source at certain node.
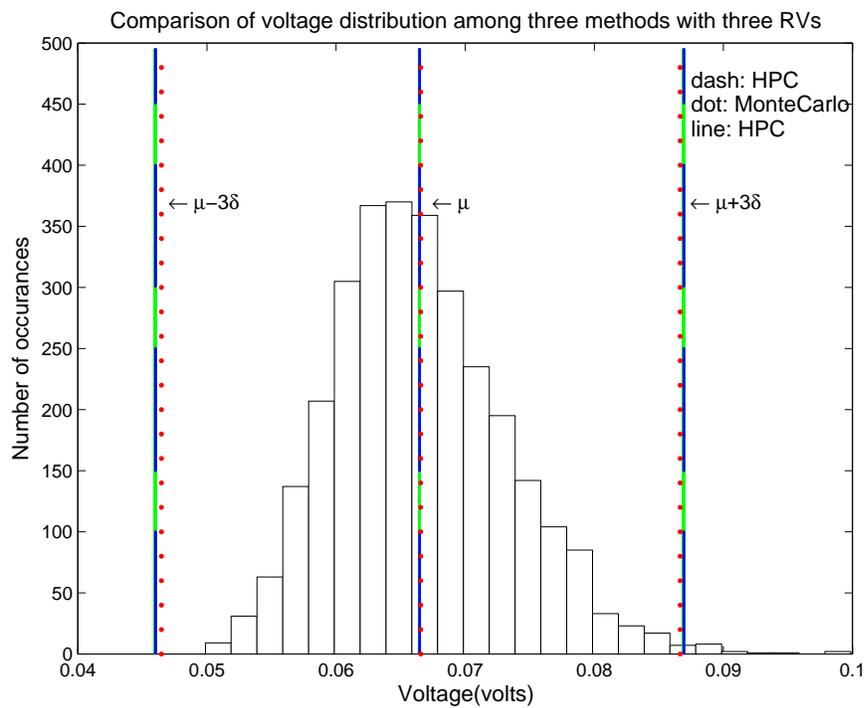
Figure 5.9: Distribution of the voltage variations in a given node by StoEKS, HPC and Monte Carlo of a circuit with 280 nodes with three random variables using the time-invariant leakage model. $g_i = 0.1 I_p$

5.5   Summary

In this chapter, we have proposed a fast stochastic method for analyzing the voltage drop variations of on-chip power grid networks. The new method, called *StoEKS*, applies Hermite polynomial chaos (HPC) to represent the random variables in both power grid networks and input leakage currents with log-normal distribution. This HPC method transforms a stochastic analysis problem into a deterministic analysis problem where increased augmented circuit matrices are created. The augmented circuit matrices consist of the coefficients of Hermite polynomials representing both variational parameters in circuit matrices and input sources. We then applied the extended Krylov subspace method (EKS) to compute variational responses from the augmented circuit equations. The proposed method does not require any sampling operations as used by collocation based spectral stochastic analysis method. Experimental results have shown that the proposed method is about two-order magnitude faster than the existing Hermite PC based simulation method and more order of magnitudes faster than Monte Carlo method with marginal errors. StoEKS also increases the analysis capacity of pervious statistical simulation methods based on the spectral stochastic method.

Chapter 6

Non-linear Variational
Interconnect Delay calculation

6.1 Problem Statement

The process-induced variability has huge impacts on the circuit timing in the sub-100nm VLSI technologies [40, 8]. Process variations include variations at different levels (wafer level, die-level and within a die) and they are caused by different sources (lithograph, materials, aging etc) [12]. Some of the variations are systematic like those caused by lithography process. Some are purely random like the doping density of impurities and edge roughness.

Process variational impacts on the interconnect circuit delay have to be characterized for robust statistical static timing analysis (SSTA). The state of art SSTA engines now assume non-linear delay models from both the interconnect and gates in the so-called canonical form to ensure the accuracy of SSTA [11, 31, 8]. As a result,

close form expressions for delays of interconnects in terms of variable parameters are required.

Existing works on the statistical interconnect modeling and analysis works, however, mainly focus on deriving the probability density function (PDF) or accumulated distribution function (CDF) of interconnect delays. One simple way to do this is by means of Monte Carlo (MC) based sampling method, which can be applied to existing delay modeling method such as Elmore delay, first time moment [27], and D2M based on the first two moments [6]. Monte Carlo method is most flexible and trusted method. However, its high computing costs render its applications limited to very small circuits. Also MC can't generate closed form delay expressions in terms of variable parameters, required by the SSTA.

The closed form delay model for interconnect was proposed in [5], which, however, only assumes linear and Gaussian delay models for interconnect. On the other hand, many fast statistical interconnect modeling techniques have been proposed in the past to compute the PDF/CDF of delays [34, 15, 60, 33, 37, 22].

Method in [34] is based on the model order reduction (MOR) with the Taylor series representation (first-order and second order) of the variations. Authors in [15, 33] applied multi-dimensional MOR, where moments with respect to complex frequency variables s and random variables are computed. But its computation costs grow very rapidly with the increasing number of variables as the number of moments grow exponentially. Interval-valued statistical modeling and model order reduction methods were proposed in [37, 36]. Interval methods in general suffer the over-pessimism problem in spite of the recent improvement by using affine interval arithmetic. Also, the errors are accumulated with the arithmetic operations.

Recently statistical interconnect analysis and modeling based on the orthogonal polynomial analysis was proposed [60, 59, 22], where statistical variations are presented by orthogonal polynomial representation (OPR) and Galerkin approach is used to obtain the coefficients. The major benefit of this method is that it can convert a statistical problem into a deterministic problem as one only needs to solve for the coefficients of the polynomials deterministically in order to compute the variations of the responses or performance metrics. The authors in [60, 59] first applied the Hermite polynomial method to obtain the statistical voltage and delaying information. The Hermite-polynomial method, however, can result in very large circuit matrices (called augmented matrices) and the matrix sizes grows with the increase of random variable count. Also the method can't generate delays in the OPR forms, which are needed for the SSTA. This problem is partially mitigated by [22], which compute the statistical moment in Hermite polynomial form. But the final delay expressions are only represented by those moments, not the variable parameters.

In this chapter, we propose a general approach to generate the non-linear interconnect analytic delay models in terms of variable parameters. The new delay models are represented by the OPR form, which can easily be converted to many existing delay canonical forms required by the SSTA methods. Our approach is different from the existing orthogonal polynomial based method [60, 22] in the following aspects: First the new approach computes the analytic delay models (the coefficients of orthogonal polynomials) directly using the efficient numerical quadrature method. While [60, 59] computes the closed form expression only for voltage response first and then MC-like method is used to compute the CDF/PDF of the delays. While method [22] only computes closed form expressions in terms of variational moments in the OPR form.

Experimental results show that the proposed delay models is very accurate compared with MC simulation in estimating statistical delay of interconnects.

6.2 New method for Variational Interconnect Delay model

In this section, we present the proposed new method, which is based on efficient numerical Gaussian quadrature and spare grid quadrature to compute the analytic variational interconnect delays. Once the delay in the form of Hermite polynomial is obtained, the variational characteristic of the delay, such as mean, standard deviation can further be computed.

6.2.1 New algorithm flowchart

First, we present new algorithm, *QuadDelay*, which is shown in Fig. 6.1. We first generate Gaussian quadrature point $\Theta_1^2 = \{\gamma_1, \gamma_2, \gamma_3\}$ set and weight $w = \{w_1, w_2, w_3\}$ set for one dimension of level two (for order two Hermite polynomials) in step 1. Based on one-dimension points and weights, in step 2, we then generate the Smolyak spare grid quadrature point and weight sets to get n dimension Gaussian quadratures point $\Theta_n^2 = \{\gamma_1, \gamma_2, ..., \gamma_l\}$ set and weight $w = \{w_1, w_2, ..., w_l\}$ set of level two, considering the number of n random variables in conductance $G(\xi)$ and capacitors $C(\xi)$, where l is the size of Θ_n^2.

For each given quadrature point, we compute the delay, $d(\gamma_i)$ of a specific node in step 4 and step 5. To speed up the delay computation, we perform the reduction on the interconnect system first and then perform the transient simulation to compute

the delay for the given input $u(t)$, which can be step or ramp function.

Finally, when the delays $d(\gamma_1),...,d(\gamma_l)$ at all the quadrature points are computed, we can precede, step 7, to compute the Hermite polynomials coefficients for the final analytic delay expression, $d(\xi)$, in terms of random variables ξ.

Algorithms: QUADDELAY

Input: variational interconnect system $G(\xi)$, $C(\xi)$, B, $u(t)$

Output: the HPC coefficients of delays of specific nodes, $d(xi)$

---

1. Generate the one-dimensional quadrature point set of second order Θ_1^2 and weight set w.

2. Generate the n-dimensional Smolyak quadrature point sets of second order Θ_n^2 and corresponding weight set w_n.

3. For i $= 1$ to size(Θ_n^2)

4.   Perform the PRIMA-like reduction on $G(\gamma_i)$, $C(\gamma_i)$, B to get reduced model $\mathcal{G}(\gamma_i)$, $\mathcal{C}(\gamma_i)$, $\mathcal{B}$.

5.   Calculate the delay $d(\gamma_i)$ at the current quadrature point

6. end

7. Compute the coefficients of Hermite polynomials for the delay $d(\xi)$

Figure 6.1: The proposed QuadDelay algorithm.

We remark that we are to obtain better accuracy by using the higher order orthogonal polynomials. But we find that the second order Hermite polynomials are

accurate enough for the delay calculation.

In the following, we elaborate some important steps in the proposed method.

6.2.2 Gaussian quadrature technique

The Gaussian quadrature method is an efficient numerical method to compute the definite integral of a function [25]. We apply it to compute the coefficients $a_k(t)$ in Eq.3.11 as:

$$a_k(t) = \frac{< v(t,\xi), H_k(\xi) >}{< H_k^2(\xi) >} \tag{6.1}$$

We review the method based on the Hermite polynomial below.

Our goal is to compute the integral equation $< x(\xi), H_j(\xi) >$ numerically. In this case, this problem boils down to the one-dimensional numerical quadrature problem based on the Hermite polynomials [49]. Specifically, for Hermite polynomials, we have

$$< x(\xi), H_k(\xi) > \quad = \quad \frac{1}{\sqrt{(2\pi)}} \int x(\xi) H_k(\xi) e^{-\frac{1}{2}\xi^2} d\xi \tag{6.2}$$

$$\approx \quad \sum_{i=0}^{P} x(\xi_i) H_i(\xi_i) w_i \tag{6.3}$$

Here, $\xi = \{\xi\}$, contains only one random variable. ξ_i and w_i are Gaussian Hermite quadrature abscissas (quadrature points) and weights.

The Quadrature rule basically says that if we select the roots of Pth Hermite Polynomial as the quadrature points, the quadrature is exact for all polynomials of degree $2P - 1$ or less for Eq.6.2. This is called (P-1)-level accuracy of Gaussian-Hermite quadrature here.

For multiple random variables, which require multi dimensional quadrature. The

traditional way to compute the multi dimensional quadrature is to use a direct tensor product based on one dimensional Gaussian Hermite quadrature abscissas and weights [42]. With this method, the number of quadrature points need for n dimension (variables) and Pth level is about $(P + 1)^n$, which is well known as the curse-of-dimensionality.

6.2.3 Smolyak quadrature for multi-dimensional integration

Smolyak quadrature [42] is used as an efficient method to reduce the number of quadrature points (also called sparse grid quadrature). Let's define one-dimensional sparse grid quadrature point set $\Theta_1^P = \{\gamma_1, \gamma_2, ..., \gamma_P\}$, which uses P points to achieve degree $2P + 1$ of exactness. The level-P sparse grid for n-dimensional quadrature chooses points from the following set:

$$\Theta_n^P = \bigcup_{P+1\leq|\vec{i}|\leq P+n} (\Theta_1^{i_1} \times ... \times \Theta_1^{i_n}) \tag{6.4}$$

where $|\vec{i}| = \sum_{j=1}^{n} i_j$. And the corresponding weight is:

$$w_{j_{i_1}...j_{i_n}}^{i_1...i_n} = (-1)^{P+n-|\vec{i}|} \begin{pmatrix} n - 1 \\ n + P - |\vec{i}| \end{pmatrix} \prod_m w_{j_{im}}^{i_m} \tag{6.5}$$

where $\begin{pmatrix} n - 1 \\ n + P - |\vec{i}| \end{pmatrix}$ is a combination number and w is the weight for the corresponding quadrature points. It was shown that interpolation on a Smolyak grid

ensures an error bound for the mean-square error [42]

$$|E_P| = O(N_P^r (log N_P)^{(r+1)(n-1)}), \tag{6.6}$$

where N_P is the number of quadrature points and r is the order of the maximum derivative that exists for the delay function. The number of quadrature points increases as $O(\frac{n^P}{(P)!})$.

We further prove following Proposition 1 as our theoretical basis for choosing the sparse grid level.

Proposition 1 *Sparse grid of at least level P is required for an order P representation.*

Proof Sketch. The approximation contains order P polynomials for both $x(\xi)$ and $H_j(\xi)$ for some j, so there exists $x(\xi)H_j(\xi)$ with order $2P$, which requires sparse grid of at least level P with degree $2P + 1$ of exactness.

Therefore, level 2 and level 1 sparse grid are required for quadratic and linear model, respectively. The number of quadrature points is about $2n$ and $2n^2$ for the linear and the quadratic model respectively. The time cost is about the same as the Taylor-conversion method, while keeping the accuracy of homogenous chaos expansion.

In addition to the sparse grid technique, we also employs the several accelerating techniques summarized as follows:

- When n is too small, the number of quadrature points for sparse grid may be larger than that of direct tensor product of Gaussian quadrature. For example, if there are only 2 variables, the number is 5 and 14 for level 1 and 2 sparse

grid, compared to 4 and 9 for direct tensor product. In this case, the sparse grid will not be used.

- The set of quadrature points (6.4) may contain the same points with different weights. For example, the level 2 sparse grid for 3 variables contain 4 instances of the point (0,0,0). Combining these points by summing the weights reduces 3 times of computation of $x(\vec{\gamma}_i)$.

6.2.4 Gaussian quadrature method for variational delay calculation

In this section, we discuss how to apply the Gaussian quadrature to compute the variational delay based on the Hermite polynomials.

In our method, we assume that the wire width, W, and the wire thickness or the pitch between wire, T are Gaussian variables. As a result, both the wire resistors and capacitors are no longer Gaussian as $R \propto 1/TW$ and $C \propto W/T$. The wire delay, which approximately is RC, it is no longer Gaussian either.

In our work, we use the following second-order Hermite polynomials to present the conductor and capacitor matrix, to model the distributions of conductor and capacitor:

$$G(\xi) = G_0 + G_1\xi_W + G_2\xi_T + G_3\xi_W\xi_T$$

$$C(\xi) = C_0 + C_1\xi_W + C_2\xi_T + C_3\xi_W\xi_T \tag{6.7}$$

For n random variable in conductance $G(\xi)$ and capacitors $C(\xi)$, we then generate the Smolyak quadrature points based on the method presented in subsection 6.2.3.

The selected set of point γ_i here is

$$\gamma_i = \{\gamma_1^{i_1}, \gamma_1^{i_2}, ..., \gamma_1^{i_n}], \gamma_i \in \Theta_n^2 \tag{6.8}$$

where $\gamma_1^{i_1}$ is a scalar value and a quadrature point γ_i is a vector. Each quadrature point is associated with a weight computed in Eq.6.5.

After obtaining the quadrature points Θ_n^2, we solve for the delay at each quadrature point to compute the final analytic delay expression. We assume that the final delay analytic expression is written as

$$
\begin{aligned}
d(\xi) &= \sum_{k=0}^{N} t_k H_k^2(\xi) \\
&= t_0 + \sum_{k=1}^{n} t_k \xi_k + \sum_{k=1}^{n} t_{n+k}(\xi_k^2 - 1) + \\
&\quad \sum_{k=1}^{n-1} \sum_{j=k}^{n} t_{2n+k+j} \xi_k \xi_j
\end{aligned}
\tag{6.9}
$$

where $\xi = [\xi_1, \xi_2, ..., \xi_n]$. Our task here is to compute the Hermite polynomial coefficients $t_i, i = 1, ..., P$. Each t_k in Eq.6.9 will be calculated as

$$t_k = \frac{< d(\xi), H_i(\xi) >}{< H_i^2(\xi) >} = \frac{\sum_{i=0}^{M} d(\gamma_i) H_k(\gamma_i) w_i}{< H_i^2(\xi) >} \tag{6.10}$$

Here, $d(\gamma_i)$ is the delay computed at the quadrature point γ_i.

We need to compute the delays at all the quadrature point first. Then we can use Eq.6.10 to compute the Hermite coefficients for the final delay expression. We remark that the proposed method can use any deterministic delay computation method. In our implementation, we first perform the reduction on the original networks and then

solve the reduced network in the time domain to compute the delay for the given input $u(t)$ (step or ramp).

With the Hermite Polynomial coefficients of delay, we can calculate the mean, standard deviation. We can also get the distribution of delay $d(\xi)$ by the characteristic of Hermite PC coefficients and the distribution of $\xi = \{\xi_1, \xi_2, ..., \xi_n\}$.

6.3   Experiments

All the proposed methods have been implemented in Matlab 7.0. All the experimental results are carried out on a Linux workstation with quad-core Intel Xeon CPUs with 2.99Ghz and 16GB memory.

6.3.1   Statistical delay analysis

Fig. 6.2 and Fig. 6.3 show the comparison of variational delay distributions of PDF and CDF forms, respectively. The variations in width and thickness are of 10% and 10%. The variation of second order is $10\% \times 10\% = 1\%$. The variation for capacitors is 5%, 5% and the variation of second order is $5\% \times 5\% = 0.25\%$.

The interconnect circuits is in the size of 100. From the two figures, we can see that the PDF and CDF results of QuadDelay fit the curves of Monte Carlo very well.

Table 6.1 shows the CPU running time comparison between our proposed method and Monte Carlo of 5000 times. The #*nodes* represents the number of node in the testing interconnect in all. The speedup is the running time of our proposed method over Monte Carlo with 5000 times. When the interconnect size is large, the model order reduction method, PRIMA, is used to cut the computation complexity. From
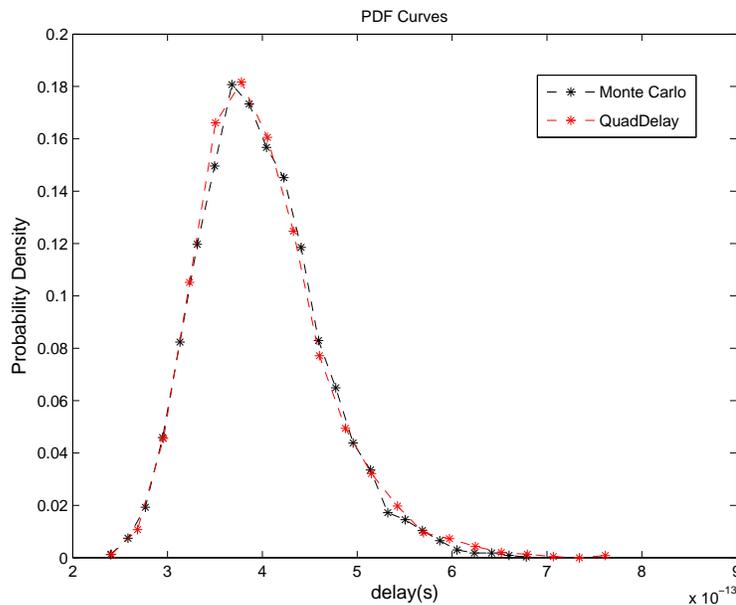
Figure 6.2: Comparison of the probability density functions (PDF) of delays between QuadDelay and Monte Carlo

the table we can see that our proposed approach is about 100 times faster than Monte Carlo. When the interconnect is large, it will take too long time to use Monte Carlo. While QuadDelay takes reasonable time to calculate the delay.

6.3.2 QuadDelay with first and second order comparison

In this subsection, we show it is necessary to use the second order orthogonal polynomials for the distribution of conductance and capacitors. To this line, we show the accuracy differences by using first order and second order representations of delay in QuadDelay.

Fig. 6.4 shows the comparison between Monte Carlo method with 5000 times and our proposed method with sparse grid quadrature. The size of the circuit is 1089 and
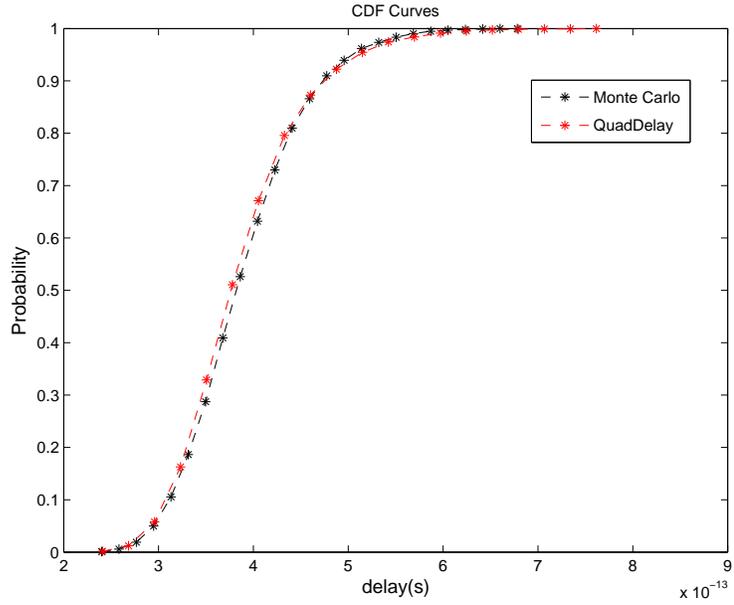
Figure 6.3: Comparison of the cumulative density functions (CDF) of delays between QuadDelay and Monte Carlo

#nodes	PRIMA	QuadDelay	MC	speedup
4	N	0.15	47.10	94.2
100	N	1.92	219.17	114.2
1089	Y	4.79	695.5	145.2
14900	Y	131.99	12994.9	98.5
74800	Y	1119.01	—	—
110889	Y	1778.46	—	—

Table 6.1: CPU time comparison of QuadDelay with Monte Carlo method.

model order reduction method is used here to reduce delay computation complexity. The solid lines are the mean and standard deviation (s.d.), which are calculated by MC. The dashed lines are the mean and standard deviation for the second order expansion for QuadDelay. The dotted lines are the mean and standard deviation of the first order expansion. From the figure we can see that the second order expansion of QuadDelay has less error than the first order expansion.
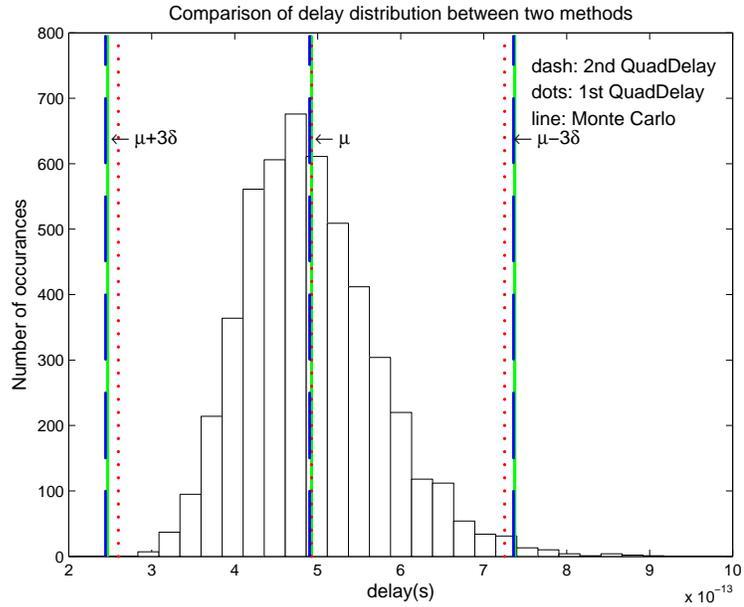
Figure 6.4: Variational delay comparison between 1st and 2nd order QuadDelay against the Monte Carlo method.

Table 6.2 and table 6.3 are the accuracy and error comparison of first and second order expansion of our proposed method, QuadDelay, over Monte Carlo. From these two tables, we can see that, the error of first order expansion is much larger than that of second order expansion. It is necessary to expand delay as Hermite Polynomials for second order.

# nodes	Mean			Std Dev		
	MC $\times 10^{-13}$	2nd $\times 10^{-13}$	1st $\times 10^{-13}$	MC $\times 10^{-13}$	2nd $\times 10^{-13}$	1st $\times 10^{-13}$
4	0.82	0.81	0.81	0.13	0.13	0.12
100	3.98	3.97	3.97	6.41	6.40	6.15
1089	4.92	4.90	4.92	8.18	8.19	7.76
14900	1966	1967	1966	315.2	316.6	304.4

Table 6.2: Accuracy comparison of QuadDelay and Monte Carlo.

# nodes	2nd % Error in μ	1st % Error in μ	2nd % Error in σ	1st % Error in σ
4	0.38	0.36	1.66	5.30
100	0.13	0.18	0.11	4.03
1089	0.45	0.02	0.21	5.15
14900	0.09	0.02	0.43	3.43

Table 6.3: Error comparison of QuadDelay between the first order and second order expansions against the Monte Carlo method.

6.4   Summary

In this chapter, we have proposed an efficient variational delay computation algorithm, which is based on orthogonal polynomial representation (OPR) of statistical variations to deal with non-linear non-Gaussian interconnect delay. We can calculate the delay in form of OPR when taking in the process variation also in OPR form. We calculated transient response (under step or ramp input) to get the variational delay on corresponding points with corresponding weights obtained by efficient Gaussian Hermite numerical quadrature method and Sparse Grid acceleration. Experimental results show that the proposed method are very accurate comparing with Monte-Carlo based sampling in estimation statistical delay in more efficient way.

Chapter 7

Conclusion

This chapter concludes the dissertation and the research work about process variation analysis.

With the scaling down of VLSI semiconductor technology to 45nm, 32nm, the process variation analysis plays more important role in the integrated circuit design procedure. Worst case has to be taken into consideration for design. Most of the time, Monte Carlo is used to get the characteristic of the process variation influence on the designed chips. In this thesis, some statistical methods are introduced to do process variational analysis for power grid network analysis and variational interconnect delay analysis.

Spectral statistical analysis for variational power grid network is proposed first. The process variation in power grid network includes leakage current input variation, which is considered as lognormal distribution, interconnect wire width and length variation, which is considered as Gaussian distribution in extracted resistors and capacitors. Hermite Polynomial Choas are used here, based on which MNA matrix G, C

and input u are expanded in Fourier-like polynomial. The statistical problem is transformed to deterministic problem then. The experimental results show that it can deal with random process with large fluctuation. While using Taylor expansion considering large variation, the error becomes large comparing with Monte Carlo. Also, when spatial correlation is considered in intra-die variation, principal component analysis (PCA) is used here to transform dependent random variables into independent ones first. The experimental results show the necessity of the usage of principal component analysis.

However, when the number of variables need to consider increases, the augmented matrix would grow large faster. The approach *StoEKS* is presented as one of the solution. First, Hermite Polynomial Choas are used to transform the statistical problem to a deterministic problem using Galerkin method. A deterministic augmented matrix is obtained. It then uses extended Krylov subspace(EKS) to reduce the size of the augmented matrix. *StoEKS* is able to take care of more random variable with larger size circuit. Experimental results show it is either two magnitude faster than only using Hermite Polynomial Choas or more faster than Monte Carlo, only with marginal error.

All above are solving statistical linear process variation. Here, *QuadDelay* is proposed for nonlinear and linear, non-Gaussian interconnect delay. The input and output for the interconnect delay calculation are all orthogonal polynomial representation of statistical variations. The calculation process includes the obtaining of sampling points and corresponding points weights. The points and their weights are obtained by Gaussian numerical quadrature method. The number of sampling points is based on the number of points and the level of accuracy need to achieve. Still,

the number is much less than the number of Monte Carlo samplings. *QuadDelay* is efficient and as accurate as Monte Carlo. How to solve the problem with large number of random variables efficiently still remains to be a problem. How to do the sampling, which means to make less sampling points to achieve tolerant accurate still has room to develop. Also, process variation considering non-linear and non-Gaussian random variables need to be taken care of too.

Our research can also be extended. Nowadays, how to simulate large scale power grid network is still a problem. This problem should also be solved with taking process variations into consideration. How to calculate worst case under process variation correctly and efficiently still need to be studied. Recently, parrellel computing, including multicore and multithread, are used for large scale computation. How to use use more threads in one computer or more computers can be investigated for process variation analysis for power grid or interconnect.

Bibliography

[1] Umfpack. http://www.cise.ufl.edu/research/sparse/umfpack/.

[2] International technology roadmap for semiconductors(itrs) 2005, 2006 update, 2006. http://public.itrs.net.

[3] G. Adomian. Stochastic green's functions. *Proceedings of Symposia in Mathematical Physics and Engineering*, 16:1–39, 1964.

[4] G. Adomian and K. Malakian. Inversion of stochastic partial differential operators - the linear case. *J. Math. Anal. Appl.*, 77:309–327, 1980.

[5] K. Agarwal, D. Sylvester, D. Blaauw, F. Liu, S. Nassif, and S. Vrudhula. Variational delay metrics for interconnect timing analysis. In *Proc. Design Automation Conf. (DAC)*, pages 381–384, 2004.

[6] C. J. Alpert, A. Devgan, and C. Kashyap. RC delay metrics for performance optimization. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 20(5):571–582, May 2001.

[7] A. T BharruchaReid. *Probabilistic Methods in Applied Mathematics*. Academic press, 1968.

[8] D. Blaauw, K. Chopra, A. Srivastava, and L. Scheffer. Statistical timing analysis: From basic principles to state of the art. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 27(4):589–607, APR. 2008.

[9] Y. Cao, Y. Lee, T. Chen, and C. C. Chen. HiPRIME: hierarchical and passivity reserved interconnect macromodeling engine for RLKC power delivery. In *Proc. Design Automation Conf. (DAC)*, pages 379–384, 2002.

[10] H. Chang and S. Sapatnekar. Statistical timing analysis under spatial correlations. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(9):1467–1482, Sept. 2005.

[11] H. Chang and S. S. Sapatnekar. Full-chip analysis of leakage power under process variations, including spatial correlations. In *Proc. Design Automation Conf. (DAC)*, pages 523–528, New York, NY, USA, 2005. ACM.

[12] C. Chiang and J. Kawa. *Design for Manufacturability*. Springer, 2007.

[13] E. Chiprout. Fast flip-chip power grid analysis via locality and grid shells. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 485–488, Nov. 2004.

[14] Courant and Hilbert. *Methods of Mathematical Physics*. Interscience, New York, 1953.

[15] L. Daniel, O. C. Siong, L. S. Chay, K. H. Lee, and J. White. Multi-parameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 23(5):678–693, May 2004.

[16] V. De and S. Borkar. Technology and design challenges for low power and high performance. In *Proc. Int. Symp. on Low Power Electronics and Design(ISLPED)*, pages 163–168, Aug. 1999.

[17] J. Fan, S. X.-D. Tan, Y. Cai, and X. Hong. Partitioning-based decap capacitor budgeting via sequence of linear programming. *Integration, the VLSI Journal*, 40(4):516–524, Mar. 2007.

[18] I. A. Ferzli and F. N. Najm. Statistical estimation of leakage-induced power grid voltage drop considering within-die process variations. In *Proc. Design Automation Conf. (DAC)*, pages 865–859, 2003.

[19] I. A. Ferzli and F. N. Najm. Statistical verification of power grids considering process-induced leakage current variations. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 770–777, 2003.

[20] R. Ghanem. The nonlinear Gaussian spectrum of log-normal stochastic processes and variables. *Journal of Applied Mechanics*, 66:964–973, December 1999.

[21] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Dover Publications, 2003.

[22] P. Ghanta and S. Vrudhula. Variational interconnect delay metrics for statistical timing analysis. In *Proc. Int. Symposium. on Quality Electronic Design (ISQED)*, pages 19–24, 2006.

[23] P. Ghanta, S. Vrudhula, R. Panda, and J. Wang. Stochastic power grid analysis considering process variations. In *Proc. European Design and Test Conf. (DATE)*, volume 2, pages 964–969, 2005.

[24] W. Guo, S. X.-D. Tan, Z. Luo, and X. Hong. Partial random walks for transient analysis of large power distribution networks. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Science(IEICE)*, E87A(12):3265–3271, Dec 2004.

[25] A. Iserles. *A First Course in the Numerical Analysis of Differential Equations*. Cambridge University, 3 edition, 1996.

[26] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.

[27] A. Kahng and S. Muddu. Two pole analysis of interconnection trees. In *IEEE Multi-Chip Module Conference*, pages 105–110, Feb. 1995.

[28] T. Karnik, S. Borkar, and V. De. Sub-90nm technologies-challenges and opportunities for CAD. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 203–206, San Jose, CA, Nov 2002.

[29] J. N. Kozhaya, S. R. Nassif, , and F. N. Najm. A multigrid-like technique for power grid analysis. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 21(10):1148–1160, Oct. 2002.

[30] Y. Satish Kumar, Jun Li, Claudio Talarico, and Janet Wang. A probabilistic collocation method based statistical gate delay model considering process variations and multiple input switching. In *DATE '05: Proceedings of the conference on Design, Automation and Test in Europe*, pages 770–775, Washington, DC, USA, 2005. IEEE Computer Society.

[31] J. Xiong L. Cheng and L. He. Non-linear statistical static timing analysis for non-gaussian variation sources. In *Proc. Design Automation Conf. (DAC)*, pages 250–255, 2007.

[32] Y. Lee, Y. Cao, T. Chen, J. Wang, and C. Chen. HiPRIME: Hierarchical and passivity preserved interconnect macromodeling engine for rlkc power delivery. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(6):797–806, 2005.

[33] Xin Li, Peng Li, and L. Pileggi. Parameterized interconnect order reduction with explicit-and-implicit multi-parameter moment matching for inter/intra-die variations. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 806–812, 2005.

[34] Ying Liu, Lawrence T. Pileggi, and Andrzej J. Strojwas. Model order-reduction of rc(l) interconnect including variational analysis. In *DAC '99: Proceedings of the 36th ACM/IEEE conference on Design automation*, pages 201–206, 1999.

[35] M. Loeve. *Probability theory*. Springer Verlog, 4 edition, 1977.

[36] J. D. Ma and R. A. Rutenbar. Fast interval-valued statistical modeling of interconnect and effective capacitance. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 25(4):710–724, April 2006.

[37] James D. Ma and Rob A. Rutenbar. Fast interval-valued statistical interconnect modeling and reduction. In *ISPD '05: Proceedings of the 2005 international symposium on Physical design*, pages 159–166, 2005.

[38] N. Mi, J. Fan, and S. X-D Tan. Analysis of power grid networks considering lognormal leakage current variations with spatial correlation. In *Proc. IEEE Int. Conf. on Computer Design (ICCD)*, pages 56–62, 2006.

[39] N. Mi, J. Fan, and S. X-D Tan. Simulation of power grid networks considering wires and lognormal leakage current variations. In *Proc. IEEE International Workshop on Behavioral Modeling and Simulation (BMAS)*, pages 73–78, 2006.

[40] S. Nassif. Delay variability: sources, impact and trends. In *Proc. IEEE Int. Solid-State Circuits Conf.*, pages 368–369, San Francisco, CA, Feb 2000.

[41] S. Nassif. Design for variability in DSM technologies. In *Proc. Int. Symposium. on Quality Electronic Design (ISQED)*, pages 451–454, San Jose, CA, Mar 2000.

[42] E. Novak and K. Ritter. Simple cubature formulas with high polynomial exactness. *Constructive Approximation*, 15(4):449–522, Dec 1999.

[43] A. Odabasioglu, M. Celik, and L.T. Pileggi. PRIMA: Passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pages 645–654, 1998.

[44] M. Orshansky, S. R. Nassif, and D. Boning. *Design for Manufacturability and Statistical Design*. Springer, 2007.

[45] S. Pant, D. Blaauw, V. Zolotov, S. Sundareswaran, and R. Panda. A stochastic approach to power grid analysis. In *Proc. Design Automation Conf. (DAC)*, pages 171–176, 2004.

[46] J. R. Phillips and L. M. Silveira. Poor man's TBR: a simple model reduction scheme. In *Proc. European Design and Test Conf. (DATE)*, pages 938–943, 2004.

[47] J. R. Phillips and L. M. Silveira. Poor man's TBR: a simple model reduction scheme. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(1):43– 55, 2005.

[48] L. T. Pillage and R. A. Rohrer. Asymptotic waveform evaluation for timing analysis. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, pages 352–366, April 1990.

[49] W. H. Press, S. A. Teukolsky, W. T. Veterling, and B. P. Flannery. *Numerical Recipes in C: The art of Scientific Computing*. Cambridge University Press, 1992.

[50] R. Rao, A. Srivastava, D. Blaauw, and D. Sylvester. Statistical analysis of subthreshold leakage current for VLSI circuits. *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, 1(2):131–139, Feb 2004.

[51] C. Koh S. Zhao, K. Roy. Decoupling capacitance allocation and its application to power-supply noise-aware floorplanning. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 21(1):81–92, Jan. 2002.

[52] L. Scheffer, L. Lavagno, and G. Martin. *EDA for IC Implementation, Circuit Design, and Process Technology*. Taylor & Francis Group,LLC, 2006.

[53] J. Shi, Y. Cai, J. Fan, S. X.-D. Tan, and X. Hong. Pattern-based iterative method for extreme large power/ground analysis. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 26(4):680–692, Apr. 2007.

[54] M. Shinozuka and T. Nomoto. Response variability due to spatial randomness of material properties. In *Technical Report*, 1980.

[55] A. Srivastava, R. Bai, D. Blaauw, and D. Sylvester. Modeling and analysis of leakage power considering within-die process variations. In *Proc. Int. Symp. on Low Power Electronics and Design(ISLPED)*, pages 64–67, Aug. 2002.

[56] B. G. Streetman and S. Banerjee. *Solid-State Electronic Devices*. Prentice Hall, 2000. 5th ed.

[57] X.-D. Tan and C.-J. Shi. Efficient very large scale integration power/ground network sizing based on equivalent circuit modeling. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 22(3):277–284, March 2003.

[58] X.-D. Tan, C.-J. Shi, D. Lungeanu, and J.-C. Lee. Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programmings. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 22(12):1678–1684, Dec. 2003.

[59] S. Vrudhula, J. M. Wang, and P. Ghanta. Hermite polynomial based interconnect analysis in the presence of process variations. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 25(10), 2006.

[60] J. Wang, P. Ghanta, and S. Vrudhula. Stochastic analysis of interconnect performance in the presence of process variations. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 880–886, Nov 2004.

[61] J. M. Wang and T. V. Nguyen. Extended Krylov subspace method for reduced order analysis of linear circuit with multiple sources. In *Proc. Design Automation Conf. (DAC)*, pages 247–252, 2000.

[62] K. Wang and M. Marek-Sadowska. On-chip power supply network optimization using multigrid-based technique. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 24(3):407–417, Mar. 2005.

[63] D. Xiu and G.Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *J. of Computational Physics*, (187):137–167, 2003.

[64] Yamazaki, M. Shinozuka, and G. Dasgupta. Neumann expansion for stochastic finite element analysis. In *Technical Report*, 1985.

[65] 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*, 21(2):159–168, Feb. 2002.

[66] Y. Zou, Y. Cai, Q. Zhou, X. Hong, S. Tan, and L. Kang. Practical implementation of stochastic parameterized model order reduction via hermite polynomial chaos. In *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, pages 367–372, 2007.