REDUNDANCY-AWARE ELECTROMIGRATION CHECKING FOR MESH POWER GRIDS

by

Sandeep Chatterjee

A thesis submitted in conformity with the requirements for the degree of Master of Applied Sciences
Graduate Department of Electrical & Computer Engineering
University of Toronto

© Copyright 2013 by Sandeep Chatterjee
Abstract

Redundancy-Aware Electromigration Checking for Mesh Power Grids

Sandeep Chatterjee

Master of Applied Sciences
Graduate Department of Electrical & Computer Engineering
University of Toronto
2013

Electromigration (EM) is re-emerging as a significant problem in modern integrated circuits (IC). Especially in power grids, due to shrinking wire widths and increasing current densities, there is little or no margin left between the predicted EM stress and that allowed by the EM design rules. Statistical Electromigration Budgeting (SEB) estimates the reliability of the grid by considering it entirely as a series system. However, a power grid with its many parallel paths has much inherent redundancy. In this work, we propose a new model to estimate the MTF and reliability of the power grid under the influence of EM, which accounts for these redundancies. We refer to this as the mesh model. To implement the mesh model, we also develop a framework to estimate the change in statistics of an interconnect as its effective-EM current varies. The proposed algorithm is quite fast and has an overall observed empirical complexity of $O(n^{1.5-1.6})$. The results indicate that the series model, which is currently used in the industry, gives a pessimistic estimate of power grid MTF and reliability by a factor of 3-4.
Acknowledgements

First and foremost, I would like to thank Professor Farid N. Najm for his insightful recommendations, continuous guidance and inspiration, without which this work would not have been possible. He is the best supervisor one could ever hope for. I cherish the weekly meetings with him, in which we discussed and analyzed the progress made in my research. His excellent technical expertise and mentorship is something I aspire for. Thank you Professor, for giving me the opportunity to work with you; it has been an amazing experience that lead to my holistic development as a professional, a researcher and as a person.

I would also like to thank Professor Andreas Veneris, Professor Jianwen Zhu and Professor Tony Chan Carusone from the ECE department at the University of Toronto for reviewing this work and for providing me with constructive comments. I would also like to acknowledge the financial support provided by the University of Toronto and Natural Sciences and Engineering Research Council (NSERC) of Canada.

I am grateful to Mohammad B. Fawaz for his friendship and support. I thank him for all the good times we have had together, inside as well as outside the lab. The work presented in chapter 3 is a collaborative effort with him, and I enjoyed the fruitful discussions we had while working together.

Many thanks to Abhishek, who was my go-to-person when I was new to Toronto. I am thankful for his support during the first year of my degree program. He was always there to answer my questions on almost any topic related to my research and otherwise. I would also like to take this opportunity to thank my colleagues for providing a great and pleasant environment. I am especially grateful to Noha Sinno, Zahi Moudalla, Zissis Poulos, Jason Luu, Fitsum Assamnew, Braiden Brousseau and Jongsook Choi. I wish them best in their endeavors.

I am also lucky to have the support of my friends Aakash Nigam, Hari Sridhar, Anmol Ratna Singh, Nitish Srivastava and many others in Toronto with whom I spent quality
time outside my work. I will never forget the numerous Toronto adventures or the late night discussions on random topics I had with them. Also, many thanks to my friends in India for their support and motivation. I wish you all the best.

My biggest gratitude goes to my parents, Mr. Jitendra Kr. Chatterjee and Mrs. Soma Chatterjee for always encouraging me and wishing only the best for me. They never doubted my ability and invested in my future, at times compromising their present. Thank you both for all the support, this work is dedicated to you.

Lastly, I offer my regards to those whom I might have missed, but supported me in any respect during the completion of this work.
# Contents

1 Introduction ................................................. 1
   1.1 Motivation ............................................. 1
   1.2 Contribution ............................................ 3
   1.3 Organization ........................................... 3

2 Background .................................................. 5
   2.1 Introduction ............................................. 5
   2.2 Electromigration ......................................... 5
      2.2.1 Electromigration Lifetime Model: Black’s equation .................... 6
      2.2.2 Blech Effect ......................................... 7
      2.2.3 Effective-EM Current .................................. 8
   2.3 EM Reliability Mathematics .............................. 8
      2.3.1 Basic Concepts ....................................... 9
      2.3.2 EM Time-to-Failure Distributions ........................... 10
      2.3.3 Normal Distribution ................................... 12
      2.3.4 Lognormal Distribution ................................. 14
      2.3.5 TTF sample Generation ................................ 15
   2.4 Power Grid Model ........................................... 15
   2.5 Monte Carlo random sampling approach ................... 17
      2.5.1 Mean Estimation ...................................... 18
4.4.2 Estimating grid MTF ............................................. 55
4.5 Experimental Results ............................................. 56

5 Reliability Estimation and Selective TTF updates 67
  5.1 Introduction ....................................................... 67
  5.2 Survival Probability and Reliability Estimation ................. 67
  5.3 Selective TTF updates for speed-up ............................. 70
    5.3.1 Motivation .................................................. 70
    5.3.2 Overview of approach ...................................... 70
  5.4 Parameter Variation ............................................. 72
  5.5 Implementation and Experimental Results ....................... 74

6 Conclusions and Future Work 82

Appendices 84

A The Banachiewicz-schur form 85

Bibliography 86
List of Tables

3.1 Comparison between Woodbury Formula Solution Update and Sparse-LU using UMFPACK ........................................... 37

4.1 Details of Power Grids used in experiments ......................... 58

4.2 Comparision of Power Grid MTF $\mu_s$ and $\mu_m^*$ as obtained using the Series model and Mesh model for the original approach. $\mu_s$ and $\mu_m^*$ are estimated with 95% confidence and maximum 5% relative error. ..................... 60

4.3 Speed and accuracy of the secondary approach, $\mu'_m$ is estimated with 95% confidence and maximum 5% relative error. ............. 61

5.1 Survival probability Estimation of power grids for user-provided values of $y$ with 95% confidence and maximum absolute error of $\pm 0.05$. ............... 76

5.2 Speed and accuracy after using selective TTF update approach to estimate power grid MTF, $|\pi(R)|$ is the average size of the set $\pi(R)$ (in terms of % of interconnects) over all failures when $\delta^*_v = 10^{-3}V$ and $\epsilon^+$ is the percentage error in MTF estimation with respect to the original approach. .... 78
List of Figures

1.1 Nanometer technology trends [2] .................................................. 2

2.1 A typical triple point in a wire ..................................................... 6

2.2 Typical $f(t)$ plot for the normal and the lognormal distributions .... 13

2.3 A typical DC model of power grid .............................................. 16

2.4 Definition of $z_{\alpha/2}$ .............................................................. 19

2.5 Comparison of Statistical Bounds for 95% and 99% confidence, where

   Eq. A: $\left(\frac{z_{\alpha/2}}{2\varepsilon}\right)^2$, Eq. B: $\left(\frac{\sqrt{63} + z_{\alpha/2}}{2\sqrt{\varepsilon}}\right)^2$ and
   Eq. C: $\left(\frac{z_{\alpha/2}\sqrt{2\varepsilon} + 0.1 + \sqrt{(\varepsilon + 0.1)z_{\alpha/2}^2 + 3\varepsilon}}{2\varepsilon}\right)^2$ .......... 22

2.6 A typical on-die modern power grid with a mesh structure .............. 23

3.1 Resistance and void shape evolution with time [42] ....................... 27

3.2 Conductance change with EM degradation for the step-increase linear
model and open-circuit model ...................................................... 27

3.3 Flow-chart for the mesh model .................................................. 29

3.4 Comparison of average time taken to update the voltage drops ......... 38

3.5 Comparison of cumulative time taken to update the voltage drop after $p^{th}$
interconnect fails for grid IBMPG4 .............................................. 39

3.6 Empirically determined complexity of (a) Sparse-LU (UMFPACK) (b)
Woodbury and Woodbury-Banachiewicz-Schur form ......................... 39
3.7 Voltage-drop maps of IBMPG5 grid (a) at $t = 0$, when it is fresh and (b) after the first interconnect failure. As per series model, (b) has failed, but clearly, the voltage drops at all nodes are below $v_{th} = 0.18V$.

3.8 Voltage-drop maps of IBMPG6 grid (a) at $t = 0$, when it is fresh and (b) after the first interconnect failure. As per series model, (b) has failed, but clearly, the voltage drops at all nodes are below $v_{th} = 0.18V$.

4.1 Proposed approach for single-step case, (a) Current density profile of $S_0$, (b) Current density profile of $S_1$, (c) Known CDF $F_1(t)$, CDF $F_0(t)$ unknown for $t > t_1$, (d) The proposed solution.

4.2 Proposed approach for multi-step case, (a) CDF plot of $S_0$ as per the proposed approach, the square markers locate the time of current change when a new shifted lognormal is used to describe the statistics, (b) Current density profile.

4.3 Proposed approach for multi-step case, with EM-immune time-spans in between (a) CDF plot of $S_0$ as per the proposed approach (b) Current density profile.

4.4 Updating a TTF sample (a) Obtaining a TTF sample at $t = 0$ (b) Updating TTF sample when current density changes, the square marker locates the time of current-density change.

4.5 Comparison of MTF as obtained using the series model and mesh model for the original approach.

4.6 Dependence of relative error between $\mu'_m$ and $\mu'_m$ on $(v_{th} - \max(v_0))$ and $\sigma_{in}$ for IBMPG2 grid.

4.7 Average time-breakup between node voltage updates and TTF updates of surviving interconnects after an interconnect failure.

4.8 Scalability of the original approach.
4.9 Voltage drop maps of IBMPGNEW2 (a) at $t = 0$ and (b) at the time of failure grid failure. For this particular realization, 68 interconnect failures were required to exceed threshold voltage $v_{th} = 0.18V$.

4.10 Voltage drop maps of IBMPG5 (a) at $t = 0$ and (b) at the time of failure grid failure. For this particular realization, 20 interconnect failures were required to exceed threshold voltage $v_{th} = 0.18V$.

4.11 Voltage drop maps of IBMPG6 (a) at $t = 0$ and (b) at the time of failure grid failure. For this particular realization, 36 interconnect failures were required to exceed threshold voltage $v_{th} = 0.18V$.

5.1 Change in node voltage drops for a randomly chosen single interconnect failure in power grids IBMPG5, IBMPG6, IBMPGNEW2 and G4. Nodes where change in voltage drop is less than $10^{-3}$ are omitted in the figure.

5.2 Dependence of boundary of $\pi(R)$ on $\delta^*_v$, $s = \partial v_k / \partial R$.

5.3 Change in gain ratio with current exponent ($n$) and standard deviation of lognormal ($\sigma_{ln}$) for power grid IBMPG2.

5.4 Variation of gain ratio with property of metal used for power grid IBMPG2.

5.5 Variation of gain ratio with power grid configuration for power grid IBMPG2.

5.6 Estimated statistics of RVs $T_s$ and $T_m$ for IBMPG5, they are found to have a normal distribution.

5.7 Scalability of the selective TTF update approach.

5.8 Runtime ($t_{cpu}$) and error ($|\varepsilon^+|$) vs. $\delta^*_v$; effect of varying $\delta^*_v$ on $\bar{T_f}$ (average time/failure) and $|\bar{\pi}(R)|$ (the average cardinality of the set $\pi(R)$). All the results are for IBMPG6.
Chapter 1

Introduction

1.1 Motivation

Verifying the power grid is a crucial step in VLSI design, as the reliability of the underlying logic heavily depends on its power grid. Not only must an IC perform as desired, it must also survive and function as intended for several years before failing. This concern is addressed by a variety of design measures, for example the choice of materials, the length and width of interconnects and so on. The density of on-chip interconnects has increased from generation to generation of modern integrated circuits. This requires a decrease in both interconnect width and thickness and, consequently, the operating current densities increase. As such, new factors come to bear, which reduce reliability and which previously were negligible. Specifically, electromigration (EM) has re-emerged as a significant problem in modern chip design and there are three problems that demand attention: 1) existing EM checking techniques for the power grid are overly pessimistic (because of an underlying series system assumption, as will be explained later), leading to loss of safety margins and multiple design iterations, 2) increased current density in grid metal lines has led to a significant loss of margins between the predicted EM stress and the allowed thresholds, and 3) checking modern, large power grids for EM has become very expensive. To make things worse, it is forecast [1] [2] that metal line current density and reliability due to EM will get dramatically worse with continued technology scaling. As a result, EM signoff has become increasingly difficult and designers are forced to reconsider traditional approaches, and to look with suspicion at the large safety margins and pessimism built into traditional EM checking methods.

Historically, ‘worst case’ current density limits for individual lines were used to arrive at reliable designs for ICs. However, since it is never the case that all the lines in power grids will carry the worst case currents at the same time, the ‘worst case’ current density
approach severely restricted the design process. This motivated the need for a model to relate the reliability of individual components to the reliability of the entire system. In an early contribution [3] a series model was proposed to determine the reliability of an IC, under which a system is deemed to have failed as soon as any of its components fails. Under the series model, and with some simplifying assumptions, the failure rate of the system is the sum of failure rates of individual components. The series model was applied to the Alpha 21164 microprocessor, under the name Statistical Electromigration Budgeting (SEB) [4] and became a standard technique in many industrial CAD tools.

However, modern power grids use a mesh structure. As such, there are many paths for the current to flow from the C4 bumps to underlying logic, a characteristic we refer to as redundancy. Mesh power grids have much redundancy and are in fact closer to (but not quite) a parallel system, rather than a series system; and so have a longer lifetime than a series system. This issue has largely been ignored in EM checking tools, both in academia and industry; no industrial tool has this feature today. While SEB accounts for the fact that EM failures are statistical it does not, however, recognize the benefits of redundancy and it treats the overall metal structure as a series structure. A power grid is not necessarily failed if one of its metal line fails. Instead, in our approach, we deem a power grid to have failed only when enough lines have failed that the voltage on the grid becomes unacceptable. Our data shows that, in many cases, a grid can tolerate up to 70 or more line failures before it truly fails! The results obtained using publicly available grids from IBM, with up to 700k nodes, and internally generated grids, with up to 1 million nodes, show an increase in predicted lifetime of 3-4X compared to the existing series-system based approach. A lot of margin is therefore “left on the table” and there is room for high-impact improvements in EM verification.
1.2 Contribution

The goal of this research is to develop a more realistic power grid EM checking and budgeting method that takes the redundancy of the power grid structure into account. In this work, we make an attempt to move away from the overly pessimistic series model to a new model for estimating the mean time-to-failure of the power grid. We also connect the IR-drop with EM degradation to determine when a power grid is presumed to have failed. Specifically,

1. We develop a new model, referred to as the mesh model, that factors in the redundancy of the power grid while estimating its MTF and reliability. In implementing the mesh model, we also develop an efficient exact method to update the node voltage drops as the structure of the grid changes due to failure of interconnects.

2. We propose a novel framework to estimate the change in statistics of an interconnect as its effective-EM current density varies in steps on a time-scale comparable to the failure times of interconnects, as is the case in mesh model.

1.3 Organization

The thesis is organized as follows: Chapter 2 covers the necessary background material on electromigration and the associated reliability mathematics. It then familiarizes the reader with the power grid model used and describes the Monte-carlo random sampling method used for mean and probability estimation. Relevant prior work in electromigration checking is also reviewed in this chapter. Chapter 3 presents the mesh model in detail. Among other things, an implementation of the mesh model needs to repeatedly a) solve a large linear system to find the new node voltage drops and b) update the failure statistics of the surviving interconnects, as lines fail due to electromigration. The second part of chapter 3 tackles the issue of finding new voltage drops: it presents a highly efficient exact method based on the Woodbury formula and the Banachiewicz-Schur form to update the node voltage drops as the conductance matrix $G$ undergoes low rank updates. The second issue is dealt in chapter 4, which proposes a novel framework to update the failure statistics of the surviving interconnects. In Chapter 5, we first extend the framework developed so far to estimate the probability of survival of a power grid for a given period of time based on Monte-carlo random sampling approach. Next, we describe a heuristic approach to speed-up the MTF estimation engine by performing selective TTF updates. Finally, it is shown that the reduction in pessimism as predicted
by the mesh model is stable under parameter variation. We conclude and give future research directions in Chapter 6.
Chapter 2

Background

2.1 Introduction

In this chapter, we will review the basic background material. Section 2.2 reviews the key literature on electromigration relevant to this work. In the next section, basic probability terms essential for understanding the stochastic reliability analysis are defined. In section 2.4, we focus on the power grid model to be used. In section 2.5, we turn our focus to the Monte Carlo random sampling approach, where we derive the stopping criteria (in terms of number of samples to be obtained) for estimating the mean of a distribution or the probability of occurrence of an event. Finally, in section 2.6, we briefly review the previous electromigration checking techniques used for power grids.

2.2 Electromigration

Electromigration is the mass transport of metal due to momentum transfer between electrons (driven by an electric field) and diffusing metal atoms. Failure occurs in metal lines only when there is a flux divergence with regard to the movement of metal atoms. It commonly occurs at distortions in the lattice, in the form of vacancies and/or grain boundaries which can be schematically represented by triple points (see Figure 2.1). These allow for diffusion and/or migration of metal atoms that lead to flux divergence. Flux divergence arises when the flow of metal atoms into the region is not equal to the outflow of atoms from the region. A void is created due to depletion (when outflow from a region is greater than the flow into it) and a hillock is created due to accumulation (when the flow into a region is greater than the outflow).
2.2.1 Electromigration Lifetime Model: Black’s equation

Under the influence of Electromigration, metal line resistance increases as a line approaches failure and starts to deform due to void creation or hillock formation. Empirically, Black’s equation has been used to determine the Mean Time to Failure or MTF of a metal line under the influence of electromigration [5]:

\[ MTF = \frac{w_d t}{A} J^{-n} \exp \left( \frac{E_a}{kT_m} \right) \]  

(2.1)

where \( A \) is an empirical constant that depends on a host of physical properties such as volume resistivity of the metal and effective ionic scattering cross section for electrons, \( w_d \) and \( t \) are the width and thickness of the line respectively, \( J \) is the effective-EM current density, \( n \) is a current exponent, \( k \) is the Boltzmann’s constant, \( T_m \) is the temperature in Kelvin and \( E_a \) is the activation energy for electromigration. Since the power grid is composed of layers made up of the same metal, the physical properties used to arrive at \( A \) remain constant throughout the grid, implying that \( A \) is simply a scaling factor in (2.1). The activation energy \( E_a \) and current exponent \( n \) are empirically determined. A current density exponent \( n = 1 \) is consistent with void growth limited failure since the rate of unconstrained void growth is proportional to the current density [6][7]. An exponent \( n = 2 \) indicates void nucleation limited failure in the model based on the Korhonen analysis [8] and other theoretical works [9][10]. If \( 1 < n < 2 \), then failure is due to a combination of nucleation and growth [11] [12], but if \( n > 2 \), the failure is due to Joule heating induced temperature gradients or due to short length effects [12].

Voids may lead to open-circuits or unacceptable resistance increase in a line, whereas protrusion of hillocks usually cause short-circuits between adjacent lines and inter-level conductors at positions where conductors cross over each other or in devices where there are two spaced layers, such as a capacitor [13]. Hillocks due to EM can also result in thin dielectrics, which are further susceptible to dielectric breakdown [14]. In this work, we
assume for simplicity that all interconnect failures due to EM are caused by nucleation and/or growth of voids.

### 2.2.2 Blech Effect

Electromigration is mainly caused due to divergence of atomic flux induced by the electron wind. This atomic flux is opposed by the back-stress developed due to accumulation of atoms at the anode end. If the length of line is short enough, then this back stress could overcome the critical stress required for nucleation of a void in a line (or even stop the growth of void in a line), and thus the line is no longer susceptible to EM failure. Blech was the first to observe this phenomenon [15]. He designed an experiment where conductor islands were deposited onto a titanium nitride (TiN) film and stressed at a high current density. He noted that when the stripe was reduced to a certain length, the metal transport stopped. This observation was explained by considering the driving force of atomic flux as a combination of two forces: the electron wind force and the mechanical back-stress force. The atomic flux can then be expressed as [15] [16] [17] [18]:

$$J_a = \frac{D_a C_a}{kT}(F_{elec} + F_{mech}) = \frac{D_a C_a}{kT} \left( |Z^*|e\rho J - \Omega \frac{\partial \Lambda}{\partial x} \right) \quad (2.2)$$

where $D_a$ is atomic diffusivity, $C_a$ is atomic concentration, $\Lambda$ is the hydrostatic stress, $\Omega$ is the atomic volume, $\rho$ is the resistivity of the material, $Z^*$ is the effective atomic charge and $e$ is the fundamental atomic charge. In steady state, the electron wind force balances the mechanical back-stress force, so that $J_a = 0$, and the stress gradient is given as:

$$\frac{\partial \Lambda}{\partial x} = \frac{|Z^*|e\rho J}{\Omega} \quad (2.3)$$

Integrating (2.2) over the length of line yields:

$$\Lambda(x) = \Lambda_0 + \frac{|Z^*|e\rho J}{\Omega} x \quad (2.4)$$

where $\Lambda_0$ is the stress at $x = 0$. Equation (2.4) implies that the maximum stress-gradient developed in an interconnect is determined by its current density. Hence, if the current density changes, the stress-gradient changes.

Given that $\Lambda_{nucl}$ is the hydrostatic stress required for void nucleation, a critical product for electromigration failure can be stated as:

$$\beta_c = (JL)_c = \frac{\Omega(\Lambda_{nucl} - \Lambda_0)}{Z^*e\rho} \quad (2.5)$$
In other words, for a given interconnect that has not already failed, if the product of its current density \( J \) and length \( L \) is less than \( \beta_c \), the maximum stress developed due to the present current density is not enough for a void to nucleate, which means that the interconnect is immune to electromigration failure. This effect is referred as the Blech effect and \( \beta_c \) is the critical Blech product. For a given current density \( J \), a line is said to be EM-immune if \( JL \leq \beta_c \) and EM-susceptible if \( JL > \beta_c \).

### 2.2.3 Effective-EM Current

Because EM is a long-term cumulative failure mechanism, the changes in line current on very short time-scales (such as the normal operation of a digital chip), are not terribly significant. Instead, the standard approach is to compute an effective-EM current, which is a constant current value, derived from the line current waveform that gives the same lifetime for that line under the influence of electromigration. In traditional EM work, this effective current is computed based on some assumed periodic current waveform. If the waveform is uni-directional, then Direct Current EM analysis is used, based on time-averaged current density \( J_{avg} \) [19].

\[
J_{DC,eff} = J_{avg} = \frac{1}{T} \int_0^T J(t)dt
\]  

(2.6)

For more general cases of bi-polar currents, transient current EM analysis uses an effective-EM current density of [20]:

\[
J_{tran,eff} = \frac{1}{T} \left( \int_0^T J^+(t)dt - \phi \int_0^T |J^-(t)|dt \right)
\]  

(2.7)

where \( \int J^+(t)dt \) is the integral of one side of the current waveform and is larger than the opposite side (i.e. \( \int J^-(t)dt \)) and \( \phi \) is the EM recovery factor, that is determined experimentally.

### 2.3 EM Reliability Mathematics

The Time-to-Failure (TTF) of a line under the influence electromigration is subject to manufacturing variations. To elaborate, consider an experiment E1 with a large population set \( S \), of \( N \) conductors (lines), made of the same material, using the same manufacturing process and with identical dimensions. All the conductors are stressed with the same current density \( J \) at a temperature \( T_m \) for time \( t \geq 0 \). Then, it is found that the
TTF of all the conductors are not exactly the same. Instead, it is observed that the TTF of the conductors in $S$ are distributed over some time-range. In other words, if we pick any single conductor from the set $S$, the only thing we know is the frequency distribution of the TTFs obtained from the experiment $E_1$. The notion of random variable is used to (rigorously) deal with such scenarios: one can view the different TTFs obtained from the population as different realizations of a single conductor, whose TTF is modeled as a random variable. In this section, basic probability terms and concepts essential for understanding the stochastic reliability analysis of EM are presented.

2.3.1 Basic Concepts

The concept of a random variable, combined with the four basic probability functions Cumulative Distribution Function (CDF), Probability Distribution Function (PDF), Reliability Function and Failure/Hazard rate Function are the fundamental building blocks of reliability mathematics used to describe a non-deterministic (random) process, such as degradation due to EM.

Random Variable: A Random Variable or RV is a function that maps the outcome of a random process to a real number. In this document, they are represented by bold face capital letters. RVs can be classified as either continuous or discrete. Because the TTF of a single line can assume any value in a given time-range, it is described by a continuous RV, henceforth denoted by $T$. In formal terms, $T$ stands for the TTF of a single randomly chosen line such that the sets $\{T \leq t\}$ represent the events that the line has failed before time $t$, and have assigned probability values. An RV can be completely characterized by its PDF or CDF, so that there is no further need to refer to the underlying probability space [21].

Cumulative Distribution Function: A Cumulative Distribution Function or CDF of a RV $T$, denoted by $F(t)$ is defined as:

$$F(t) \triangleq \mathcal{P}\{T \leq t\}$$  \hspace{1cm} (2.8)

where the right hand side represents the probability that the RV $T$ takes a value less than or equal to $t$. In other words, the probability that a single randomly chosen conductor will fail by time $t$ is given by $F(t)$. $F(t)$ is a monotonically increasing function in $t$, with $F(-\infty) = 0$ and $F(\infty) = 1$. The RHS of (2.8) can be evaluated if the type of distribution, the mean and the standard deviation of $T$ are known beforehand, or if we have a sufficient number of TTF samples.
**Probability Distribution Function**: A Probability Distribution Function or PDF of a RV \( T \) is a function, denoted by \( f(t) \), defined as:

\[
f(t) \triangleq \frac{dF(t)}{dt} \tag{2.9}
\]

The probability for \( T \) to fall within a particular interval \([t_1, t_2]\) is given by the integral of the PDF over that region:

\[
\mathcal{P}\{t_1 < T \leq t_2\} = F(t_2) - F(t_1) = \int_{t_1}^{t_2} f(t) dt \tag{2.10}
\]

**Reliability function**: A Reliability Function, denoted by \( R(t) \), of a RV \( T \) is defined as:

\[
R(t) \triangleq 1 - F(t) = \mathcal{P}\{T > t\} \tag{2.11}
\]

where the right hand side represents the probability that the RV \( T \) takes a value greater than \( t \). In other words, \( R(t) \) is the probability that the single randomly chosen conductor will survive beyond time \( t \). \( R(t) \) is a monotonically decreasing function in \( t \), with \( R(-\infty) = 1 \) and \( R(\infty) = 0 \).

**Failure rate**: The Failure rate, \( \lambda(t) \) is defined as the (instantaneous) rate of failure for the survivors up to time \( t \) during the next instant of time [22]. Mathematically:

\[
\lambda(t) \triangleq \lim_{\Delta t \to 0} \frac{\text{Probability of failure in } [t, t + \Delta t], \text{ given still operational at } t}{\Delta t} = \lim_{\Delta t \to 0} \frac{\mathcal{P}\{t < T \leq t + \Delta t\}}{\Delta t \cdot \mathcal{P}\{T > t\}} = \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{f(t)}{R(t)} \tag{2.12}
\]

Thus, the failure rate, or the hazard rate, describes the conditional probability of failure around a time point \( t \). \( \lambda(t) \) is instantaneous, it may change the next instant, and the units that have already failed play no further role since only the survivors count.

### 2.3.2 EM Time-to-Failure Distributions

The type of distribution function used to characterize the lifetimes arising from a wide range of failure mechanisms (including electromigration), can be established in two ways:

1. **Theoretical approach**: In this approach, the lifetime distribution model is chosen
based on the physics of failure mechanism. This is the ideal approach, but in general, it is very hard to derive the distribution from the underlying interactions. Hence, the second approach is preferred.

2. **Empirical approach**: The frequency distribution obtained from any experiment (such as E1 on page 8), when normalized with respect to the total number of conductors $N$, gives us a (coarse) staircase approximation of the PDF. Equation (2.10) can then be used to find the CDF $F(t)$ by setting $t_1 = 0$ and $t_2 = t$. There is another easier way to find the CDF $F(t)$: define $N_{\text{failed}}(t)$ to be the number of conductors that have failed in the time-period $[0, t]$. Then, from definition, the CDF is given by:

$$F(t) \approx \frac{N_{\text{failed}}(t)}{N} \quad (2.13)$$

As $N$ is increased, the normalized frequency distribution of TTF samples approaches the actual PDF and RHS of (2.13) approaches the actual CDF. However, since $N$ cannot be increased beyond a limit, the actual practice in industry is to use goodness-of-fit methods on the experimental data to choose one of known distribution models as the lifetime distribution model for a given failure mechanism.

For electromigration, there have been many (theoretical and experimental) investigations regarding which distribution model to follow. Venables et. al. [23] present a model that relates the time to failure due to EM to the density of initial voids in the metal stripe. Using an experimentally observed distribution of this density they derive the corresponding distribution of time to failure which turned out to be very close to the lognormal distribution. Gall et. al. [24] studied the early failures due to EM on a total of 20,000 interconnects, and showed that lognormal is valid. Lloyd et. al. [25] explain the lognormal observed behavior of electromigration as the result of a normal distribution of activation energies ($E_a$) and a lognormal distribution of grain sizes. However, it is pointed out that the lognormal is not extendable (failure time of a series system of lognormal life elements is not lognormal), and the multi-lognormal (an extreme value lognormal) is more appropriate. Yoh et. al. [26] propose a shifted-lognormal (SLN) as the lifetime distribution model of individual long wires, and showed that shifted lognormal and truncated multi-lognormal are very similar. However, the lognormal still remains a practical and simple model for the Time-to-failure of long metal lines, and is used throughout this work.

Another justification for the use of lognormal as time to failure distribution model of long metal lines comes from the fact that degradation due to EM is mainly due to diffusion or migration of atoms, which is expected to follow multiplicative degradation
property. If at any instant in time a degradation process undergoes a small increase in the total amount of degradation that is proportional to the current total amount of degradation, then it is reasonable to expect the time to failure (i.e. reaching a critical amount of degradation) to follow a lognormal distribution [22]. This can be formally explained as follows:

Let \( y_1, y_2, \ldots, y_n \) be the amount of degradation for a particular failure process taken at successive discrete instants of time as the process moves towards failure. Assume that the following relationship exists between the measured quantities:

\[
y_i - y_{i-1} = \eta_i y_{i-1} \\
y_i = (1 + \eta_i)y_{i-1}
\]  

(2.14)

where \( \eta_i \) are random independent shocks to the system that move the degradation process forward; then the degradation at the \( n^{th} \) instant is given by:

\[
y_n = \left[ \prod_{i=1}^{n} (1 + \eta_i) \right] y_0
\]

(2.15)

where \( y_0 \) is a constant. Taking natural logarithms on both sides of (2.15):

\[
\ln y_n = \sum_{i=1}^{n} \ln(1 + \eta_i) + \ln y_0 \approx \sum_{i=1}^{n} \eta_i + \ln y_0
\]

(2.16)

From the Central Limit Theorem, it can be asserted that \( \sum_{i=1}^{n} \eta_i \) (and \( \ln y_n \)) has approximately a normal distribution [27], and thus by the properties of lognormal distribution (as presented in section 2.3.4), \( y_n \) has a lognormal distribution for any \( n \) (or at any continuous time \( t \)). Typically, failure happens when degradation reaches a critical predefined point, and and thus the time of failure caused by a multiplicative degradation process can be modeled by a lognormal distribution.

### 2.3.3 Normal Distribution

The normal distribution, also called the Gaussian distribution, is one of the best known and most widely used distributions. If a RV \( T \) has a normal distribution, then its PDF is bell shaped, and is given by:

\[
f(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{t-\mu}{\sigma} \right)^2}, \quad -\infty < t < +\infty
\]

(2.17)
Chapter 2. Background

Figure 2.2: Typical $f(t)$ plot for the normal and the lognormal distributions

where $\mu \equiv E[T]$ and $\sigma^2 \equiv Var(T)$ ($Var$ denotes variance). The PDF is symmetrical around $\mu$ and attains its maximum value of $\frac{1}{\sigma\sqrt{2\pi}} \approx 0.4$ at $t = \mu$. When $\mu = 0$ and $\sigma = 1$, the resulting normal distribution is called standard normal distribution, and its PDF and CDF are given by:

$$
\phi(t) = \frac{1}{2\pi} e^{-\frac{1}{2}t^2}, \quad \Phi(t) = \int_{-\infty}^{t} \frac{1}{2\pi} e^{-\frac{1}{2}t^2} dt \quad -\infty < t < +\infty \quad (2.18)
$$

If $T$ is normal with mean $\mu$ and variance $\sigma$, then the RV $Z = \frac{T - \mu}{\sigma}$ is a RV with standard normal distribution. Thus, CDF of $T$ can be denoted as:

$$
F(t) = \Phi \left( \frac{t - \mu}{\sigma} \right) \quad (2.19)
$$

There is no closed form expression for $\Phi$. However, on computers, it can be evaluated with the aid of the $\text{erf}$ and $\text{erfc}$ functions. The CDF and failure rate of the normal distribution can thus be written as:

$$
F(t) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{t - \mu}{\sigma\sqrt{2}} \right) \right] \quad (2.20)
$$

$$
\lambda(t) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{1}{2} \left( \frac{t - \mu}{\sigma} \right)^2} \text{erfc} \left( \frac{t - \mu}{\sigma\sqrt{2}} \right) \quad (2.21)
$$
2.3.4 Lognormal Distribution

A random variable \( T \) is said to have a lognormal distribution when \( \ln T \) has a normal distribution. The Probability Distribution Function of a lognormal distribution is given by:

\[
f(t) = \frac{1}{t\sigma_{\ln}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - \mu_{\ln}}{\sigma_{\ln}}\right)^2}, \quad 0 < t < +\infty
\]

where \( \mu_{\ln} \triangleq E[\ln T] \) and \( \sigma_{\ln}^2 \triangleq Var(\ln T) \). A lognormal RV \( T \) can be transformed to a standard normal variable \( Z \) where \( Z = (\ln T - \mu_{\ln})/\sigma_{\ln} \). Hence, the Cumulative Distribution Function can be written as:

\[
F(t) = \int_{-\infty}^{t} \frac{1}{x\sigma_{\ln}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu_{\ln}}{\sigma_{\ln}}\right)^2} dx
= \Phi\left(\frac{\ln t - \mu_{\ln}}{\sigma_{\ln}}\right) \quad 0 < t < +\infty
\]

where \( \Phi \) is the CDF of Standard Normal Distribution. If the distribution originates at some time \( t = \Delta \neq 0 \), then the modified PDF and CDF functions are written as:

\[
f(t - \Delta) = \frac{1}{(t - \Delta)\sigma_{\ln}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t - \Delta) - \mu_{\ln}}{\sigma_{\ln}}\right)^2} \quad \Delta < t < +\infty
\]

\[
F(t - \Delta) = \Phi\left(\frac{\ln(t - \Delta) - \mu_{\ln}}{\sigma_{\ln}}\right) \quad \Delta < t < +\infty
\]

Note that \( \Delta > 0 \) implies that a distribution originated after \( t = 0 \), while \( \Delta < 0 \) means that the distribution originated before \( t = 0 \).

Define \( \mu_T \triangleq E[T] \) and \( \sigma_T^2 \triangleq Var[T] \). The relations between \( \mu_{\ln}, \sigma_{\ln}, \mu_T \) and \( \sigma_T \) are as follows:

\[
\sigma_{\ln}^2 = \ln\left(1 + \left(\frac{\sigma_T}{\mu_T}\right)^2\right)
\]

\[
\mu_{\ln} = \ln(\mu_T) - \frac{1}{2} \ln\left(1 + \left(\frac{\sigma_T}{\mu_T}\right)^2\right) = \ln(\mu_T) - \frac{1}{2} \sigma_{\ln}^2
\]

\[
\mu_T = \exp\left(\mu_{\ln} + 0.5\sigma_{\ln}^2\right)
\]

\[
\sigma_T^2 = \mu_T^2(\exp(\sigma_{\ln}^2) - 1)
\]

In the case of electromigration, the MTF or \( \mu_T \) of the lognormal distribution is given by Black’s equation (2.1). The deviation of \( \ln T \), denoted by \( \sigma_{\ln} \), is dependent on the ratio of width to median grain size of the metal line for aluminum [28] and on the product
Chapter 2. Background

$JL^2$ for copper [29], and this dependence is determined experimentally. However, the deviation in both cases eventually levels off to a constant value. Hence, in this work, it is assumed that $\sigma_{ln}$ is constant for a given material, as is typical in the literature [30] [31].

2.3.5 TTF sample Generation

In order to generate samples for a RV $T$ that has a lognormal distribution, we note that $\ln T$ has a normal distribution, so that $Z = \frac{\ln T - \mu_{ln}}{\sigma_{ln}}$ has a standard normal distribution $\Phi$ (with mean 0 and variance 1). Hence, if $\Psi$ is a sample from $\Phi$ and $\tau$ is the corresponding sample from $T$, we have:

$$\Psi = \frac{\ln \tau - \mu_{ln}}{\sigma_{ln}}$$

$$\tau = \exp (\mu_{ln} + \Psi \sigma_{ln})$$

(2.29)

Given that the Time-To-Failure of an interconnect in a power grid is modeled with a lognormal distribution, a TTF sample from its distribution can be obtained using (2.29). The random samples $\Psi$ from $\Phi$ are obtained using the Zigguart method [32].

Eq (2.29) can be simplified further by using (2.26):

$$\tau = \exp \left( \ln (\mu_T) - 0.5 \sigma_{ln}^2 + \Psi \sigma_{ln} \right)$$

$$\tau = \mu_T \exp \left( \Psi \sigma_{ln} - 0.5 \sigma_{ln}^2 \right)$$

(2.30)

Since $\sigma_{ln}$ is constant, the TTF sample $\tau$ is directly proportional to the MTF $\mu_T$, for a given $\Psi$.

2.4 Power Grid Model

A complete model of the power grid would include parasitic resistances, capacitances and (possibly) inductances, with the underlying logic blocks modeled as non-negative transient current sources connected between a subset of the grid nodes and ground. The transient current waveforms are assumed to be periodic. However, electromigration is a long term cumulative failure mechanism, and as such the transients in the currents that happen in the logic circuitry (at a scale of $\sim 10^{-9}$s) do not impact the degradation process significantly. The standard practice, as was mentioned before, is to use an effective-EM current to model the lifetimes of interconnects in the power grid, which is a DC value.

A power grid is a linear system, with the circuit currents acting as inputs and node
Figure 2.3: A typical DC model of power grid

voltage drops as the outputs:

\[ v(t) = \mathcal{L}[i(t)] \]  

(2.31)

where \( \mathcal{L} \) is a linear operator representing the power grid. An interconnect in the power grid mostly carries uni-directional currents, so that its EM analysis (using Black’s Equation) depends on the average current only. Because the power grid is a linear system, the average current in an interconnect can be obtained directly from the circuit current averages by doing a DC analysis. Also, the average output response of a linear system can be found by subjecting it to the average of its input stimulus (this can be easily proved using the superposition and the time-invariance property of a linear system). Hence, the grid node voltage drop averages can be obtained directly from the circuit current averages by means of a DC analysis:

\[
\langle v(t) \rangle = \mathcal{L}[(i(t))] = \mathcal{L}_{DC}[(i(t))] 
\]  

(2.32)

In Chapter 3, we will propose a framework for EM verification that depends on user-provided thresholds on the average voltage drops on the grid nodes. In this framework, it becomes sufficient therefore to perform DC analysis of the grid, driven by the averages of the circuit currents. Therefore, a DC model of the power grid, devoid of capacitances and inductances, is sufficient. Moreover, a DC model is highly computationally advantageous.

The power grid nodes are numbered 1, 2, \ldots, m, with the ground node being 0. Let \( i \) be the vector of DC current stimuli. We assume that \( i \) is well-defined for all nodes 1, 2, \ldots, m,
so that a node \( k \) with no connected current source will have \( i_k = 0 \). Let \( u_k \) be the voltage at some node \( k \), and \( u(t) \) be the vector of all \( u_k(t) \) signals. Applying Kirchoff’s current law (KCL) at every node leads to the following matrix formulation:

\[
G(t)u(t) = -i + G_v(t)u_{dd}
\]  

(2.33)

where \( G(t) \) and \( G_v(t) \) are \( m \times m \) conductance matrices resulting from application of the traditional modified nodal analysis formulation [33]. In this case, \( G(t) \) and \( G_v(t) \) vary over large time-scales, as the lines age and deform, hence the time dependence. The matrix \( G(t) \) is known to be a diagonally dominant symmetric positive definite \( M \)-matrix (so that \( G(t)^{-1} \geq 0 \) [34]. \( G_v(t) \) is a \( m \times m \) matrix of conductance elements connected to \( v_{dd} \) sources [35], and \( u_{dd} \) is a constant vector each entry of which is equal to \( v_{dd} \). Notice that if we set \( i = 0 \), then obviously \( u_k(t) = v_{dd} \forall t \), so that the above system becomes:

\[
G(t)u_{dd} = G_v(t)u_{dd}
\]  

(2.34)

By replacing \( G_v(t)u_{dd} \) by \( G(t)u_{dd} \) in (2.33), it can be re-written as:

\[
G(t)[u_{dd} - u(t)] = i
\]  

(2.35)

We can express (2.35) in a more convenient way, directly in terms of voltage drops: define \( v_k(t) = v_{dd} - u_k(t) \) to be the voltage drop at node \( k \), and let \( v(t) \) be the vector of voltage drops, then the system equation can be written as:

\[
G(t)v(t) = i
\]  

(2.36)

This is a revised system equation which we can use directly to solve for voltage drop values. Comparing (2.36) to (2.33), it is easy to see that the circuit described by this equation consists of the original power grid, but with all voltage sources set to zero (short circuit) and all current source directions reversed.

### 2.5 Monte Carlo random sampling approach

Random sampling refers to iteratively selecting/generating sample values from the domain of a given distribution (according to their probabilities) and computing the arithmetic average of these samples as an estimate of the required quantity, a so-called Monte Carlo approach. The Monte Carlo approach can be used to estimate the mean of a distri-
bution or the probability of occurrence of an event; we use it later to estimate the power grid MTF and survival probability. The key feature of the Monte Carlo approach, as we will shortly see, is that the desired accuracy (in terms of relative error) and confidence (a measure of how certain we are about the estimate) can be specified up-front by the user.

2.5.1 Mean Estimation

Estimating the mean of a distribution is a classic problem in statistics. In this subsection, we will review a standard approach employed to achieve the same. Let \( X_1, X_2, \ldots, X_w \) be \( w \) independent RVs that are identically distributed with the same distribution \( f(x) \), that has a mean \( \mu \) and variance \( \sigma^2 \geq 0 \). Such a collection is called a random sample. The arithmetic average of this random sample is another RV \( \bar{X}_w \) given by:

\[
\bar{X}_w = \frac{X_1 + X_2 + \ldots + X_w}{w} \tag{2.37}
\]

It can be easily shown that mean of \( E[\bar{X}_w] = \mu \) and \( Var(\bar{X}_w) = \sigma^2/w \). Thus, the distribution of \( \bar{X}_w \) is narrower than \( X_i \) and tightens around its mean as \( w \) increases. In the limit, this is expressed as the law of large numbers:

\[
\lim_{w \to \infty} \bar{X}_w = \mu \tag{2.38}
\]

Sampling from a Normal distribution

Now, suppose that \( X \) has a normal distribution. In this case, it can be shown that \( \bar{X}_w \) also has a normal distribution and \( Z = \frac{\bar{X}_w - \mu}{\sigma/\sqrt{w}} \) is an RV with standard normal distribution \( \Phi \). For \( \alpha \in [0, 1] \), let \( z_{\alpha/2} \) be such that:

\[
\mathcal{P}\{Z > z_{\alpha/2}\} = 1 - \Phi(z_{\alpha/2}) = \alpha/2
\]

\[
\Phi(z_{\alpha/2}) = \mathcal{P}\{Z \leq z_{\alpha/2}\} = 1 - \alpha/2 \tag{2.39}
\]

The value of \( z_{\alpha/2} \) for a given \( \alpha \) can be easily found from statistical tables or using the \( \text{erf}(\cdot) \) function on linux systems. Due to symmetry of the standard normal distribution
Figure 2.4: Definition of $z_{\alpha/2}$

If we choose $\alpha = 0.05$, then $Z_{0.025} = 1.96$ and we have:

$$\mathcal{P} \left\{ \bar{X}_w - 1.96 \frac{\sigma}{\sqrt{w}} \leq \mu \leq \bar{X}_w + 1.96 \frac{\sigma}{\sqrt{w}} \right\} = 0.95 \quad (2.41)$$

so that, if $\bar{X}_w$ is the observed value of $\bar{X}_w$ (i.e. mean of the observed sample values $X_1, X_2 \ldots X_w$), we can assert with 95% confidence that the true mean $\mu$ lies within $1.96\sigma/\sqrt{w}$ of the observed mean:

$$\bar{X}_w - 1.96 \frac{\sigma}{\sqrt{w}} \leq \mu \leq \bar{X}_w + 1.96 \frac{\sigma}{\sqrt{w}} \quad (2.42)$$
Stopping Criterion when variance is known

To derive a stopping criterion in terms of number of samples \( w \) required, we need an upper bound on the relative error. From the previous analysis:

\[
\frac{|\bar{X}_w - \mu|}{\sigma/\sqrt{w}} \leq z_{\alpha/2} \Rightarrow \frac{|\bar{X}_w - \mu|}{|\bar{X}_w|} \leq \frac{z_{\alpha/2}\sigma}{|\bar{X}_w|\sqrt{w}} \tag{2.43}
\]

If \( w \) is large enough, so that

\[
\frac{z_{\alpha/2}\sigma}{|\bar{X}_w|\sqrt{w}} \leq \delta \tag{2.44}
\]

for some \( \delta \in (0, 1) \), then \( \delta \) places an upper bound on the relative error with \((1-\alpha)\times100\%\) confidence. Thus we have:

\[
\frac{|\bar{X}_w - \mu|}{|\bar{X}_w|} \leq \delta \quad \text{or} \quad \frac{|\bar{X}_w - \mu|}{|\mu|} \leq \frac{\delta}{1-\delta} = \epsilon \tag{2.45}
\]

where \( 0<\epsilon<1 \) is a small positive number with \( \delta = \epsilon/(1+\epsilon) \) and we assume that \( \mu \neq 0 \). The user specifies \( \epsilon \), an upper bound on the relative error with respect to the true mean \( \mu \). The stopping criterion can now be written as (from (2.44)):

\[
w \geq \left( \frac{z_{\alpha/2}\sigma}{|\bar{X}_w|\epsilon/(1+\epsilon)} \right)^2 \tag{2.46}
\]

Stopping criterion when variance is unknown

In most cases, the variance of the distribution will be unknown. Let \( s_w \geq 0 \) be the sample standard deviation, i.e.

\[
s_w^2 = \frac{1}{w-1} \sum_{i=1}^{w} (X_i - \bar{X}_w)^2 = \frac{w(\sum X_i^2) - (\sum X_i)^2}{w(w-1)} \tag{2.47}
\]

It can be shown that \( s_w^2 \) is an unbiased estimator of the variance. Then, the RV \( t = [(\bar{X}_w - \mu)/(s_w/\sqrt{w})] \) has Student’s t-distribution with \((w-1)\) degrees of freedom. For \( w \geq 30 \), values of \( t_{\alpha/2} \) become very close to \( z_{\alpha/2} \), so that standard normal provides a good approximation to the t-distribution of size 30 or more [27]. Conservatively, we use \( w \geq 50 \) so that

\[
w \geq \left( \frac{z_{\alpha/2}s_w}{|\bar{X}_w|\epsilon/(1+\epsilon)} \right)^2 \quad \text{for} \ w \geq 50 \tag{2.48}
\]
Stopping criterion when not sampling from a *Normal* distribution

In the preceding section, we derived the stopping criteria assuming a *normal* distribution for $X$. However, in general, the distribution of $X$ may not be *normal*. In such a case, we use the *central limit theorem*, which states that for large $w$, the RV $Z = \frac{\bar{X}_w - \mu}{\sigma/\sqrt{w}}$ approaches the standard normal distribution $\Phi$ (in practice, for $w$ as small as 25 or 30, distribution of $Z$ approaches $\Phi$). Further, studies [27] have shown that even when we replace $\sigma$ by its unbiased estimate $s_w$ (from (2.47)), $Z$ is fairly close to a t-distribution provided $X$ is not too skewed; and if it is, we can use a non-linear transform (such as exponential, logarithm etc.) to make the distribution more bell-shaped.

In conclusion, for most practical cases, we simply use (2.48) to determine the number of samples $w$ required in order to estimate the mean.

### 2.5.2 Probability Estimation

The *Monte Carlo random sampling approach* can also be used to estimate the probability of occurrence of an event. Later in this work, we deal with the problem of estimating the survival probability of a power grid for a period of $y$ years as it undergoes degradation due to electromigration. Suppose there is a function that takes as input a power grid and generates a sample grid Time-to-Failure. In order to determine the required survival probability, *Bernoulli trials* are performed: $w$ grid TTF samples are generated using the given function. Generation of a sample TTF is referred as a *trial*. For each TTF sample, we have a *success* if the grid survives up to $y$ years, else we have a *failure*. If $p$ is the probability of a success, and if $x$ successes are observed in a sequence of $w$ trials, then by the law of large numbers [27]:

$$\lim_{w \to \infty} \frac{x}{w} = p$$

(2.49)

This would be a straightforward approach to determine the survival probability of a power grid, but it suffers from the defect that there is no way to tell when $x/w$ is close enough to $p$, for the process to be stopped. The solution for the same is derived in [36], and is based on statistical estimation of proportions. The stopping criterion is found to be:

$$w = \max \left[ \left( \frac{z_{\alpha/2}}{2\mathcal{E}} \right)^2, \left( \frac{\sqrt{63} + z_{\alpha/2}}{2\sqrt{\mathcal{E}}} \right)^2, \left( \frac{z_{\alpha/2}\sqrt{2\mathcal{E}} + 0.1 + \sqrt{(\mathcal{E} + 0.1)z_{\alpha/2}^2 + 3\mathcal{E}}}{2\mathcal{E}} \right)^2 \right]$$

(2.50)
2.6 Electromigration checking in power grids

Many empirical and analytical models have been proposed to explain degradation due to EM in a metal line \([5][37][9]\). All these models represent the atomic flux divergence as some function of the current density in the line. This lead to the use of current density limits for individual interconnects to arrive at reliable designs for ICs, the idea being that if each wire is made safe from EM degradation by keeping its current density below the design limit, the chip is guaranteed to be safe. The design limits were based on ‘worst case’ estimates of the current density that is expected under use conditions, usually maximum temperature \([38]\). The designers used a very simple criterion for elec-
tromigration checking, they compared interconnect dc effective current per unit width, $i_{eff}$, to a conservative fixed limit:

$$ S = \frac{\text{Actual } i_{eff}}{\text{Design Limit } i_{eff}} $$

(2.51)

A reliable design with regards to electromigration was obtained when $S < 1$, and any interconnect with $S > 1$ was re-designed. However, this approach was very restrictive, because it is highly improbable that all the lines in power grids will carry the ‘design limit $i_{eff}$’ currents at the same time. Even if it happens, a chip dissipating power in kilowatts will be required, which is infeasible.

As the designs became more complex, there was a need to move away from this highly restrictive and pessimistic EM analysis and to elevate electromigration checking to the realm of full-chip analysis. Models and methods, such as those in BERT [39], were developed to estimate damage due to EM, but they were tested only on small test circuits. Frost and Poole [3] suggested a series model to determine the reliability of an integrated circuit. Under a series model assumption, a system is said to have failed as soon as any of its components fails. However, it was pointed out later [4] that when the total statistical risk of the chip is of concern, the explicit limit placed on each and every interconnect becomes arbitrary; it is not necessary to guarantee that each and every interconnect should have $S < 1$ for the chip to be safe. As per the series model (with some simplifying assumptions), the failure rate of the system is the sum of failure rates of individual components. This allowed budgeting the chip level reliability among the
interconnects, thus minimizing to an extent the performance limitations imposed by EM reliability goal in the design. The series model with budgeting was applied to the Alpha 21164 microprocessor, under the name *Statistical Electromigration Budgeting* (SEB) [4] and became a standard technique for EM checking in many industrial CAD tools. [38] also used the series system model and ‘reliability budgeting’ to perform full chip power grid and signal net electromigration and joule heating analysis.

Though the power grid lines are usually wider than the signal lines, they are more susceptible to EM failures because they carry uni-directional currents and thus do not experience the benefit of healing [40][41]. Added to that, the series system assumption, though valid for signal lines (which have no redundancy, except for redundant vias), is highly pessimistic for modern mesh power grids. A typical mesh power grid is shown in figure 2.6. In this work, we overcome this limitation by accounting for redundancy while estimating the failure statistics of the power grid.
Chapter 3

Mesh model for Power Grid Reliability Verification

3.1 Introduction

In this chapter, we propose a new model for on-die power grid reliability verification under the influence of electromigration. The new model takes into account the inherent redundancy of the power grid structure while estimating its Mean Time to Failure (MTF) and reliability, which is completely ignored by the present state-of-the-art EM checking tools. This new model is referred as the mesh model. Among other things, an implementation of the mesh model requires us to update the node voltage drops as the power grid suffers damage due to electromigration. To do the same, we also describe a highly efficient exact method based on Woodbury formula and Banachiewicz-Schur form to update the solution of a linear system $Gv = i$ as $G$ undergoes low rank updates.

The chapter is organized as follows. Section 3.2 introduces a simple and conservative open-circuit failure model for an interconnect in the power grid. In section 3.3, we describe a new criteria to determine power grid failure, that accounts for the redundancies in the mesh structure. Next, the mesh model is presented in detail, followed by the proposed method to efficiently update the voltage drops. The implementation and results in this chapter focus on the speed-up obtained in updating voltage drops by using Woodbury formula and Banachiewicz-Schur form over Sparse-LU.
3.2 Failure model for a single interconnect

Due to high current densities, the resistance of interconnects in the power grid changes as they undergo degradation due to EM. Resistance evolution of a typical interconnect is shown in figure 3.1; it remains almost equal to the initial value up to some time $t = \tau$ (while the void nucleates and grows), after which it experiences a step increase in its resistance value (when the void spans the entire cross-section) followed by a steep-linear increase (corresponding to the longitudinal void growth) [42]. This is referred to as the step-increase linear model of EM degradation in an interconnect. Failure of an interconnect in this model is based on a percentage increase in resistance value, which usually happens around $t = \tau$.

The step-increase linear model, though accurate, is hard to simulate due to it’s non-linearity. This work instead proposes a simple and intuitive model for degradation due to EM, the open-circuit model. As the name suggests, the failure of an interconnect in this model is defined to be an open-circuit. Thus, at $t = \tau$, the interconnect is effectively removed from the power grid. The reader should note that the open-circuit model is conservative as compared to the step-increase linear model (see figure 3.2). The former model assumes an interconnect to have failed once its resistance has risen above some threshold (derived from the percentage increase). Hence, lines continue to conduct current, albeit with high resistance, after the time-to-failure ($t = \tau$) predicted by EM models. By assuming zero conductance after the predicted TTF, the open-circuit model is thus conservative.

3.3 New criteria for power grid failure

Due to scaling, supply voltages in deep sub-micron CMOS are reducing. With a reduced $v_{dd}$, even a small drop in the local supply voltage can have a significant effect on the circuit timing. As was shown in [43], if the supply voltage varies up to $12.5\%$, one can observe (by simulation) up to $2.4X$ increase in gate delay. Thus, reliable functioning of a chip is tightly coupled to the issue of voltage drop in its power supply network. Since the voltage drop of a node directly affects the delay of the underlying logic, it then clearly implies that the underlying logic will fail if the voltage drop exceeds a certain threshold value. In other words, if the node voltage drops exceed the budget, speed-critical devices will be slowed and the chip will not meet its intended performance goal. As such, a node is deemed to be safe only when its voltage drop is below the threshold, otherwise it is said to have failed.
Figure 3.1: Resistance and void shape evolution with time [42]

Figure 3.2: Conductance change with EM degradation for the step-increase linear model and open-circuit model
A power grid is functional as long as it can ensure the reliable functioning of the underlying logic. Such a power grid is said to be \textit{safe}. From the previous paragraph, one can easily see that the \textit{safety} of the power grid is tied to the \textit{safety} of its nodes. Thus, the new failure criteria is:

\begin{quote}
\textit{A power grid is said to be safe iff all the nodes are safe, otherwise it is deemed to have failed.}
\end{quote}

Typically, a chip is designed with an expected voltage drop budget of between 5\% and 10\% of $v_{dd}$. The simulation and timing verification are performed while accounting for this budget [44]. Hence, for this work, the voltage drop threshold was defined to be 10\% of $v_{dd}$ for all nodes in the grid. The new failure criteria automatically captures the redundancies in the grid, because the voltage drop of a node is determined from the grid topology itself. As opposed to the series model, the voltage drop threshold based failure criteria allows the power grid to tolerate many interconnect failures before it is declared failed.

### 3.4 New grid failure model: Mesh model

As mentioned before, for the purpose of EM reliability verification, the power grid is modeled as a resistive mesh, with the underlying logic blocks being represented by the corresponding circuit current averages connected between a subset of grid nodes and ground. Any variation of source currents due to voltage drops is ignored. As per the new failure criteria, a \textit{voltage-drop threshold} value for every grid node (or a subset of grid nodes) is given, and is captured in the $m \times 1$ vector $v_{th}$. Note that $v_{th}$ is a user-provided threshold on the average voltage drops on the grid. To avoid trivial cases, it is assumed that $i \neq 0$ and $v_{th} > 0$.

At $t = 0$, the \textit{fresh} power grid satisfies the following conditions:

1. The grid is \textit{connected}, so that there is a resistive path from any node to any other node that does not go through a $v_{dd}$ supply node or a ground node.

2. The voltage drops on all the nodes of the power grid are below the specified thresholds, i.e., if $v_0 = v(0)$, then we have $v_0 < v_{th}$.

These conditions ensure that the power grid is \textit{safe} at $t = 0$.

Since EM is not a deterministic process, there are many possible sequences in which the interconnects could fail. Each sequence of interconnect failure that leads to a grid
Chapter 3. Mesh model for Power Grid Reliability Verification

Solve the power grid for voltage drops at $t = 0$

Find (and store) the MTFs for all interconnects at $t = 0$

Initialize the random number Generator $\mathcal{G}_N$

Use $\mathcal{G}_N$ to generate TTF samples for all the interconnects

Among the surviving interconnects, fail/remove the one with the lowest TTF

Update the failure statistics of all the surviving interconnects

Update the voltage drops at all the nodes

Has the power grid failed?

Update the power grid MTF estimate

Is stopping criteria satisfied?

Print the estimated MTF of the power grid

Reset the grid to its initial conditions

START

STOP

Figure 3.3: Flow-chart for the mesh model
failure, is called a realization. In any particular realization, the interconnects have assigned sample TTF values from their respective lognormal distributions. For a given realization, as time progresses the grid lines start to deform and fail due to electromigration (in increasing order of TTF samples). As per the open-circuit model, every failure corresponds to removal of an interconnect from the grid. Thus, \( \|G(t)\| \) decreases with time, which leads to an increase in \( \|v(t)\| \):

\[
\|G(t)\|\|v(t)\| \geq \|G(t)v(t)\| = \|i\| \implies \|v(t)\| \geq \frac{\|i\|}{\|G(t)\|}
\] (3.1)

In other words, as interconnects fail, the overall voltage drop on the nodes of the grid increases. Also, removing interconnects from the grid increases grid sparsity. Given that the vector of source currents \( i \) is constant, the surviving interconnects (on an average) should conduct higher currents as time progresses, which makes them more susceptible to failure due to EM. Hence, the failure statistics of the surviving lines are dependent on the previous failures. As such, the interconnects are not statistically independent in the mesh model. Thus, until the grid fails, there is a need to update the following after each interconnect failure: i) the voltage drops at the nodes of the grid and ii) the failure statistics of the surviving interconnects.

The grid is deemed to fail at the earliest time for which the condition \( v(t) \leq v_{th} \) is no longer true, which can happen either due to voltage drop(s) at some node(s) exceeding their threshold value or due to a singular grid (i.e. as the resistors are removed, a node becomes completely disconnected from the other nodes, causing \( \det(G(t)) = 0 \)).

A singular grid models the scenario where all resistors connected to a particular node have higher than threshold resistance due to EM. As a conservative approximation this node is assumed to have failed, causing the grid to fail. Once a grid has failed, it is assumed that it remains failed for all the future time. This model, used to determine the Time-to-Failure (TTF) of the power grid, is henceforth referred to as the mesh model.

In order to estimate the MTF and reliability of the power grid using the mesh model, Monte Carlo random sampling approach is employed: we simulate different realizations using the mesh model and obtain the corresponding grid TTF samples until the stopping criteria (in terms of number of grid TTF samples required) as derived in (2.48) is satisfied. The overall flow of the mesh model is shown in the flowchart of figure 3.3.
3.5 Efficiently updating Voltage Drops

As noted, in order to implement the mesh model, there is a need to update the voltage drop of the nodes and the failure statistics of the surviving interconnects after each interconnect failure. The problem of updating the failure statistics of the surviving lines is presented in Chapter 4. In this section, we describe a highly efficient approach for updating the node voltage drops in the power grid.

3.5.1 Problem definition

In order to update the voltage drops efficiently, observe that the conductance matrix $G$ changes by $\Delta G_k$ after the $k^{th}$ interconnect fails:

$$
\Delta G_k = \begin{bmatrix}
\text{row/col} & \ldots & p & \ldots & q & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
p & \ldots & -g_k & \ldots & g_k & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
q & \ldots & g_k & \ldots & -g_k & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots 
\end{bmatrix}_{m \times m}
$$

(3.2)

Note that $-\Delta G_k$ is the $m \times m$ conductance stamp of the $k^{th}$ interconnect in the MNA matrix [45]. $g_k$ is the conductance of the $k^{th}$ failed interconnect, connected between nodes $p$ and $q$ with $p < q$.

**Proposition 1.** $\Delta G_k$, as defined in (3.2), is a rank-1 matrix.

**Proof.** A rank-1 matrix has only one linearly independent column. In other words, if $\Delta G_k$ is a rank-1 matrix, all columns of $\Delta G_k$ can be expressed as a linear combination of (only) one column in $\Delta G_k$. Now, there are three unique columns in $\Delta G_k$:

$$
C^T_0 = [0 \ldots 0 \ldots 0]_{1 \times m}
$$

$$
C^T_1 = [\ldots g_k \ldots -g_k \ldots]_{1 \times m}
$$

$$
C^T_2 = [\ldots -g_k \ldots g_k \ldots]_{1 \times m}
$$

Clearly, $C_0 = 0 \times C_1$ and $C_2 = -1 \times C_1$ so that all columns of $\Delta G_k$ can be expressed in terms of $C_1$. Hence, proved. \qed

Thus, the problem of finding the new voltage drops as the interconnects are removed from the power grid can be stated as follows: Let $G_0$ be the conductance matrix at $t = 0$. 

Let $\Delta G_1, \Delta G_2, \ldots, \Delta G_k$ be the rank-1 conductance stamps of the failed interconnects and $v_0, v_1, v_2, \ldots, v_k$ be the corresponding vectors of voltage drops such that:

$$G_0v_0 = i$$

$$(G_0 + \Delta G_1)v_1 = i$$

$$(G_0 + \Delta G_1 + \Delta G_2)v_2 = i$$

$$\vdots$$

$$(G_0 + \Delta G_1 + \Delta G_2 + \ldots + \Delta G_k)v_k = i$$

It is required to find all the solution vectors $v_0, v_1, v_2, \ldots, v_k$ in an efficient way by taking advantage of the properties of $G_0$ and the low rank updates performed on it.

### 3.5.2 Woodbury Formula for voltage updates

Any $m \times m$ rank-1 matrix can be written as the outer product of two column vectors of size $m$ [46]. Specifically,

$$\Delta G_k = u_k h_k^T \text{ such that } u_k = -h_k = \sqrt{g_k}(e_q - e_p) \quad (3.3)$$

where $g_k$ is the conductance of the $k^{th}$ failed interconnect, connected between nodes $p$ and $q$ with $p < q$ and $e_\lambda$ is a column vector of size $m$ containing 1 at the $\lambda^{th}$ location and zeros at all other locations, with $e_0$ being a vector of all zeros. Define $U$ and $H$ so that after $k$ failures:

$$U \triangleq \begin{bmatrix} | & | & \cdots & | \\ u_1 & u_2 & \ldots & u_k \end{bmatrix} \quad \text{and} \quad H \triangleq -U \quad (3.4)$$

Clearly, $\sum_{j=1}^k \Delta G_j = U H^T$. Thus, the vector of voltage drops $v_k$ after $k$ interconnect failures can be written as:

$$v_k = \left( G_0 + \sum_{j=1}^k \Delta G_j \right)^{-1} i = \left( G_0 + U H^T \right)^{-1} i \quad (3.5)$$

Now, the Woodbury formula gives [47]:

$$\left( G_0 + U H^T \right)^{-1} = G_0^{-1} - \left[ G_0^{-1} U (I_k + H^T G_0^{-1} U)^{-1} H^T G_0^{-1} \right] \quad (3.6)$$
where $I_k$ is $k \times k$ identity matrix. Using (3.6) in (3.5), we have:

$$v_k = G_0^{-1} i - \left[ G_0^{-1} U(I_k + H^T G_0^{-1} U)^{-1} H^T G_0^{-1} \right] i$$

(3.7)

We know that $G_0^{-1} i = v_0$, where $v_0$ is the voltage drop vector at $t = 0$. Define

$$Z \triangleq G_0^{-1} U = [G_0^{-1} u_1 \ldots G_0^{-1} u_k]$$  

(3.8)

Clearly, both $v_0$ and $Z$ can be efficiently found using one sparse $LU$ factorization of $G_0$ (into triangular matrices $L_G$ and $U_G$) and $k + 1$ forward/backward substitutions. Hence, (3.7) can be re-written as:

$$v_k = v_0 - Z (I_k + H^T Z)^{-1} H^T v_0$$

(3.9)

Let $W_k \triangleq (I_k + H^T Z)$ and $y_k \triangleq H^T v_0$. Eq. (3.9) becomes:

$$v_k = v_0 - Z W_k^{-1} y_k$$

(3.10)

Given an initial LU factorization of $G_0$, updating the voltage drops using (3.10) will require one backward/forward substitution ($O(m^2)$), one LU solve of a dense $k \times k$ system ($O(k^3)$), matrix-vector product ($O(mk)$) and a vector-vector subtraction ($O(m)$). Thus, the overall scalability of the updates is dominated by $O(m^2)$ as long as $k << m$. Hence, by using the woodbury formula, the costly step of factoring the conductance matrix after each interconnect failure has been eliminated.

Using Woodbury formula (3.10) to update the voltage drops has two key advantages:
i) Even though the matrix $W_k$ is dense, the number of interconnects $k$ required to fail a grid is a very small fraction of the number of nodes $m$, hence finding $W_k^{-1} y_k$ by LU factorization is comparatively very cheap. ii) The matrices $Z$ and $H^T Z$ need not be calculated from scratch for each failure, but can be efficiently updated by appending appropriate vectors at the end.

However, for large grids, the number of interconnect failures required to fail a grid can be quite large. Solving a dense linear system $W_k^{-1} y_k$ using $LU$ factorization has $O(k^3)$ complexity. In effect, this means that as $k$ increases, the voltage updates using woodbury formula slows down (though it is still faster than a complete $LU$ solve). To overcome this limitation, a further refinement based on Banachiewicz-Schur form is proposed so that the complexity of solving the linear system $W_k^{-1} y_k$ is reduced to $O(k^2)$. 
3.5.3 Improving scalability using Banachiewicz-schur form

The matrix $W_k$ can be written as:

$$W_k = I_k + H^TZ = I_k - U^T G_0^{-1} U = I_k - [u_1 \ldots u_k]^T [G_0^{-1} u_1 \ldots G_0^{-1} u_k]$$

$$= \begin{bmatrix} 1 - u_1^T G_0^{-1} u_1 & -u_1^T G_0^{-1} u_2 & \ldots & -u_1^T G_0^{-1} u_k \\ -u_2^T G_0^{-1} u_1 & 1 - u_2^T G_0^{-1} u_2 & \ldots & -u_2^T G_0^{-1} u_k \\ \vdots & \vdots & \ddots & \vdots \\ -u_k^T G_0^{-1} u_1 & -u_k^T G_0^{-1} u_2 & \ldots & 1 - u_k^T G_0^{-1} u_k \end{bmatrix}$$

(3.11)

Since $G_0^{-1}$ is symmetric, then $u_i^T G_0^{-1} u_j = u_j^T G_0^{-1} u_i \ \forall i, j$. Hence, $W_k$ is symmetric and $W_k$ can be re-written in terms of $W_{k-1}$ as:

$$W_k = \begin{bmatrix} W_{k-1} & b_k \\ b_k^T & d_k \end{bmatrix}$$

(3.12)

such that

$$b_k = [-u_1^T G_0^{-1} u_k \ldots - u_{k-1}^T G_0^{-1} u_k]^T \in \mathbb{R}^{k-1}$$

$$d_k = 1 - u_k^T G_0^{-1} u_k \in \mathbb{R}$$

(3.13)

Hence, using the Banachiewicz-schur form, $W_k^{-1}$ can be expressed in terms of $W_{k-1}^{-1}$ as (see Appendix):

$$W_k^{-1} = \begin{bmatrix} W_{k-1}^{-1} + \frac{W_{k-1}^{-1} b_k b_k^T W_{k-1}^{-1}}{s_k} & -W_{k-1}^{-1} b_k \\ -b_k^T W_{k-1}^{-1} & 1 \end{bmatrix}$$

(3.14)

where $s_k$ is the Schur complement of $W_{k-1}$ in $W_k$ and can be found using:

$$s_k = d_k - b_k^T W_{k-1}^{-1} b_k$$

(3.15)

Also, after $k$ interconnect failures, $y_{k-1}$ can be updated to $y_k$ by appending $p_k \triangleq -u_k^T v_0$ at the end:

$$y_k = H^T v_0 = -U^T v_0 = -[u_1 \ldots u_k]^T v_0 = -[u_1^T v_0 \ldots u_k^T v_0]^T$$

$$= [y_{k-1}^T \ p_k]^T$$

(3.16)
Then, $W_{-1}^{-1} y_k$ can be written as:

$$W_{-1}^{-1} y_k = \begin{bmatrix}
W_{k-1}^{-1} + \frac{W_{k-1}^{-1} b_k b_k^T W_{k-1}^{-1}}{s_k} & - \frac{W_{k-1}^{-1} b_k}{s_k} \\
- \frac{b_k^T W_{k-1}^{-1}}{s_k} & 1 \\
-W_{k-1}^{-1} b_k b_k^T W_{k-1}^{-1} y_{k-1} & - \frac{W_{k-1}^{-1} b_k}{s_k} p_k \\
- \frac{b_k^T W_{k-1}^{-1} y_{k-1}}{s_k} + \frac{p_k}{s_k}
\end{bmatrix} \begin{bmatrix}
y_{k-1} \\
p_k
\end{bmatrix}$$

(3.17)

But, the previous solution $\gamma_{k-1} \triangleq W_{k-1}^{-1} y_{k-1}$ is known, therefore:

$$\gamma_k = \begin{bmatrix}
\gamma_{k-1} + \frac{W_{k-1}^{-1} b_k b_k^T \gamma_{k-1}}{s_k} - \frac{W_{k-1}^{-1} b_k}{s_k} p_k \\
- \frac{b_k^T \gamma_{k-1}}{s_k} + \frac{p_k}{s_k}
\end{bmatrix}$$

(3.18)

Define $a_k = \frac{b_k^T \gamma_{k-1} - p_k}{s_k}$. Now, rewrite (3.18) as:

$$\gamma_k = \begin{bmatrix}
\gamma_{k-1} + a_k W_{k-1}^{-1} b_k \\
-a_k
\end{bmatrix}$$

(3.19)

Hence, (3.14) and (3.19) can be used to directly update $W_{-1}^{-1}$ and $\gamma_k$ from their previous values. Notice that $W_{-1}^{-1}$ is required because, after the next interconnect failure, $W_{-1}^{-1} b_{k+1}$ is needed to compute $\gamma_{k+1}$ using (3.19). The implementation requires a single matrix-vector product ($O(k^2)$) and $O(k^2)$ additions and divisions. Hence, the complexity of solving $W_{-1}^{-1} y_k$ is reduced to $O(k^2)$. This reduction in complexity comes into play only when $k$ is high enough.

Algorithm 1 computes the new voltage drops after the $k^{th}$ interconnect fails. It also sets the grid_singular flag if the last interconnect failure made the $G$ matrix singular. When $k = 1$, $\gamma$ is easily calculated by a scalar division. For subsequent calls, $\gamma$ is updated from its previous solution using (3.19). Note that algorithm 1 implicitly receives the sequence of previous failures from matrices $Z$ and $H$. 


Algorithm 1 WB (Woodbury-Banachiewicz Voltage Update)

**Input:** $v_0, \Delta G, L_G, U_G, Z, H, W_{inv}, y, \gamma, k$

**Output:** $v_k, \text{grid\_singular}$  
\(\triangleright \text{new voltage drop vector and grid-singularity flag}\)

1: Find $u$ and $h$ s.t. $\Delta G = uh^T$ \(\triangleright \text{as shown in (3.3)}\)
2: $z \leftarrow \text{BF\_SUBSTITUTION}(L_G, U_G, u)$ \(\triangleright \text{backward-forward substitution}\)
3: Append $z$ as the $k^{th}$ column of $Z$
4: Append $h$ as the $k^{th}$ column of $H$
5: $p \leftarrow -u^Tv_0$
6: Update $y$ by appending $p$ as the $k^{th}$ element \(\triangleright \text{as per (3.16)}\)
7: if $k = 1$ then
8: \hfill $W_{inv} \leftarrow 1/(1 - u^Tz)$ \(\triangleright u^Tz \text{ is a scalar in this case}\)
9: \hfill $\gamma \leftarrow W_{inv}y$
10: else
11: \hfill Find vectors $b$ and $d$ as given in equation (3.13)
12: \hfill $W_b \leftarrow W_{inv}b$
13: \hfill $s \leftarrow d - b^TW_b$
14: \hfill if $s = 0$ then
15: \hfill \hfill $\text{grid\_singular} \leftarrow 1$ \(\triangleright \text{grid failed due to singularity}\)
16: \hfill \hfill return
17: \hfill \hfill end if
18: \hfill $a \leftarrow \frac{b^T\gamma - p}{s}$
19: \hfill $\gamma \leftarrow \gamma + a.W_b$
20: \hfill Append $-a$ as the $k^{th}$ element of $\gamma$ \(\triangleright \text{as shown in (3.19)}\)
21: \hfill Update $W_{inv}$ as given in (3.14)
22: end if
23: $v_k \leftarrow v_0 - Z.\gamma$

3.6 Experimental Results

A C++ implementation was written to test the speedup obtained by using the Woodbury-Banachiewicz-Schur form over Sparse-LU (using UMFPACK) for updating the node voltage drops of the grid as the interconnects fail. UMFPACK [48][49][50][51] is used to perform the initial LU factorization required for implementing Woodbury-Banachiewicz-Schur form. Two types of test grids were used. The first type were generated per user specifications, including grid dimensions, metal layers, pitch and width per layer. The supply voltages and current sources were randomly placed on the grid. The technology specifications were consistent with 1.1 V 65nm CMOS technology. These grids are henceforth referred to as internal grids. The second type of grids are part of the IBM power grid benchmarks [52]. These grids are dual grids (i.e. they have both $v_{dd}$ and $gnd$ rails of the power grid), but the proposed approach was tested only for $v_{dd}$ part of the grids and are referred to as external grids. Since the problem size is determined only by the node count $m$, other grid details are not given here.
A 2.6 GHz Linux machine with 24 GB of RAM was used to obtain the results. For each grid, a random failure sequence of $k$ interconnects was decided, and then the node voltage drops were updated based on the pre-determined failure sequence in separate runs using Sparse-LU, Woodbury (eq. (3.10)) and Woodbury-Banachiewicz-Schur form (eq. (3.14) and (3.19)). The results are as tabulated in Table 3.1, which reports the speed-up obtained by Woodbury and Woodbury-Banachiewicz-Schur form over Sparse-LU, and the difference in node voltage drops in terms of $\kappa_w$ and $\kappa_{wbs}$. $\kappa_w$ is defined as follows: let $v_{p,lu}$ and $v_{p,w}$ be the voltage drop vectors obtained using Sparse-LU and Woodbury formula after the $p^{th}$ interconnect failure, then:

$$\kappa_w = \max_{p \in \{1,2,...,k\}} \frac{\left\| v_{p,w} \right\| - \left\| v_{p,lu} \right\|}{m}$$

where $\left\| \cdot \right\|$ represents the sum-norm and $m$ is the number of nodes in the grid. Similarly, $\kappa_{wbs}$ is defined as:

$$\kappa_{wbs} = \max_{p \in \{1,2,...,k\}} \frac{\left\| v_{p,wbs} \right\| - \left\| v_{p,lu} \right\|}{m}$$
Figure 3.4: Comparison of average time taken to update the voltage drops

where $v_{p, wbs}$ is the voltage drop vector obtained using Woodbury-Banachiewicz-Schur formula after the $p^{th}$ interconnect failure. For each grid, $\kappa_w (\kappa_{wbs})$ gives an idea of the maximum difference in voltage drop calculation per node among all $k$ interconnect failures in case of Sparse-LU vs. Woodbury (Sparse-LU vs. Woodbury-Banachiewicz-Schur formula). From Table 3.1, we clearly observe that on an average, Woodbury formula obtains a significant speed-up of 10.93x as compared to Sparse-LU with negligible error in calculation of voltage drop per node. Woodbury-Banachiewicz-Schur formula performs slightly better and has a speed up of 11.13x, which is to be expected.

Figure 3.4 compares the average CPU-time taken by the three methods to find a new voltage drop vector. Figure 3.5 compares the cumulative time taken to update the node voltage drops up to $p^{th}$ interconnect failure for IBMPG4 grid. Lastly, figure 3.6 shows the empirically determined complexity for the three methods. Sparse-LU has the highest complexity of $O(m^{1.41})$, Woodbury has a reduced complexity of $O(m^{1.22})$ and Woodbury-Banachiewicz-Schur has a slightly improved complexity of $O(m^{1.21})$.

### 3.7 A small note on the Mesh Model

The mesh model tries to account for the redundancy in the power grid mesh structure, which provides many parallel paths for the current to flow from C4 bumps to the current sources. As mentioned before, the stress level of a grid is identified by its node voltage drop. To show that a power grid can handle more than one interconnect failure due to redundancy, we plot a voltage-
Figure 3.5: Comparison of cumulative time taken to update the voltage drop after $p^{th}$ interconnect fails for grid IBMPG4

Figure 3.6: Empirically determined complexity of (a) Sparse-LU (UMFPACK) (b) Woodbury and Woodbury-Banachiewicz-Schur form
Chapter 3. Mesh model for Power Grid Reliability Verification

Figure 3.7: Voltage-drop maps of IBMPG5 grid (a) at $t = 0$, when it is fresh and (b) after the first interconnect failure. As per series model, (b) has failed, but clearly, the voltage drops at all nodes are below $v_{th} = 0.18V$

Figure 3.7 and 3.8 show the voltage-drop maps of IBMPG5 and IBMPG6, respectively. For both the grids, $v_{th} = 0.18V$, which was not reached by failing a single interconnect. The voltage-drop maps clearly show that unless the fresh grid has high voltage drops (close to threshold $v_{th}$), the impact of a single interconnect failure is not sufficient to cause timing violations of the underlying circuitry. More detailed results on the mesh model are presented in the next chapter.
Figure 3.8: Voltage-drop maps of IBMPG6 grid (a) at $t = 0$, when it is fresh and (b) after the first interconnect failure. As per series model, (b) has failed, but clearly, the voltage drops at all nodes are below $v_{th} = 0.18V$. 
Chapter 4

Estimating EM statistics for step currents

4.1 Introduction

In the previous chapter, we introduced the mesh model, which accounted for the redundancies in the power grid while estimating its MTF. We also presented a highly efficient approach to update the node voltage drops after each interconnect failure. However, a key part was left out: an interconnect failure in the power grid changes the current-densities through all the surviving interconnects and hence affects their residual lifetime. Since the sparsity of the grid increases due to the failure of interconnects and the vector of source currents $i$ is constant, the surviving interconnects (on average) should conduct higher currents, which makes them more susceptible to failure due to EM. Thus, the failure statistics of the surviving interconnects should be modified to reflect the same. If ignored, it leads to an optimistic estimate of grid TTF, which is undesirable. In this chapter, we develop a novel approach to estimate the change in failure statistics of an interconnect when its effective-EM current density changes over time.

The chapter is organized as follows. Section 4.2 defines the problem and points out why the traditional effective-EM current model is inadequate to handle such a scenario. In section 4.3, the proposed approach is described in detail. This is followed by implementation in section 4.4 and results in section 4.5, where we compare the estimated grid MTF values as per the series and mesh model.

4.2 Problem Definition

Consider a thought experiment in which a large set $S_0$ of $N$ isolated conductors, made of the same material, using the same manufacturing process and with identical dimensions, are tested for their failure times. The testing starts at $t = 0$. Let the current densities through all the
conductors be identical and given by the following step function:

\[
J(t) = \begin{cases} 
J_0, & 0 \leq t \leq t_1 \\
J_1, & t_1 < t < \infty 
\end{cases}
\]  

(4.1)

where \(J_0 \neq J_1\) and \(t_1\) is large, such that many conductors from \(S_0\) may have failed before \(t_1\). This current profile is shown in Fig. 4.1a. One can also interpret this as a change in effective-EM current density over long time-periods. The population \(S_0\) is fresh at \(t = 0\), but as time progresses, the conductors suffer damage due to EM and start failing. We are interested in determining the RV that describes the time-to-failure statistics of the population.

Unfortunately, the effective-EM current model (2.7) is not applicable in this case because

a) It implicitly assumes that the resulting effective-EM current density is applied to all the conductors throughout their lifetime, and
b) It assumes that the transient current-waveform is periodic.

The effective-EM current model is meant to handle periodic current waveforms that vary on a much smaller time-scale as compared to the lifetime of the conductors, such as transients occurring due to logic circuit operation. Whereas, in the proposed problem definition (4.1), the current is not periodic and it changes on a time-scale that is comparable to the TTFs of the conductors. Hence, many conductors might have failed exclusively due to \(J_0\) before the current change happens. This motivates the need for a new approach to estimate the statistics of the surviving sub-population.

### 4.3 Updating Interconnect statistics

In this section, an approach to update the statistics of the TTF of a conductor as its current density steps on long time scales (comparable to its TTF) is presented. First, to motivate such an approach, the simple case of a single step change in current density as given in (4.1) is dealt with. Then, the underlying assumptions of such an approach are listed, and based on these assumptions, a more general framework to deal with the scenario of multiple step changes in current is developed. Finally, the concept of Blech length is incorporated into the proposed approach, and the relation to update the TTF of a conductor as its current-density changes is derived.

#### 4.3.1 The Single Step Case

Consider another set \(S_1\) of \(N\) conductors identical to those in \(S_0\). Suppose \(S_1\) is subjected to a current density of \(J_1\) for all \(t \geq 0\), as shown in Fig. 4.1b. Let \(F_1(t)\) be the cumulative distribution function (CDF) of the population \(S_1\) and \(F_0(t)\) be the CDF of \(S_0\). Clearly, \(F_0(t)\)
for $t \leq t_1$ and $F_1(t) \forall t$ are known to be lognormal, and $F_0(t)$ for $t > t_1$ is to be determined. Define $t'_1$ to be such that $F_1(t'_1) = F_0(t_1)$ as shown in Fig. 4.1c, where $t_1$ is the time of current change. The difference $\delta = t_1 - t'_1$ is easy to compute, as will be demonstrated later. For now, the key question is: ‘what is the failure distribution of $S_0$ after $t_1$?’ As already pointed out, traditional EM work is not helpful here. We now provide an argument to motivate a set of assumptions, which characterize the CDF of $S_0$ for $t > t_1$.

**Proposition 2.** The expected number of surviving members of a) $S_0$ at time $t_1$, and b) $S_1$ at time $t'_1$, are exactly the same.

*Proof.* The two populations $S_0$ and $S_1$ started out fresh at $t = 0$ with the same number of conductors $N$. Let $N_{s,0}$ be the expected number of surviving conductors from the set $S_0$ for $t > t_1$ and $N_{s,1}$ be the expected number of surviving conductors from the set $S_1$ for $t > t'$. We know that $F_1(t') = F_0(t_1)$. Then, from definition (2.13), we have:

$$\frac{N - N_{s,1}}{N} = \frac{N - N_{s,0}}{N} \implies N_{s,1} = N_{s,0}$$
Chapter 4. Estimating EM statistics for step currents

Hence, the expected number of surviving members of \( S_0 \) and \( S_1 \) are exactly the same at \( t_1 \) and \( t', \) respectively.

Considering the two populations a) \( S_0 \) at time \( t_1 \), and b) \( S_1 \) at time \( t'_1 \), we observe that:

1. The expected number of surviving members of the two populations are exactly the same, as proved in Proposition 2. Therefore, loosely speaking, the two populations have experienced an identical level of deterioration.

2. The two populations are subjected to exactly the same current stress \( J_1 \), as they move forward in time, i.e. \( t_1 + x \) and \( t'_1 + x \), with \( x \geq 0 \) for \( S_0 \) and \( S_1 \), respectively.

Therefore, it is expected that, going forward in time, both populations will see the same instantaneous failure rate, i.e.:

\[
\lambda_0(t_1 + x) = \lambda_1(t'_1 + x), \quad \forall x \geq 0 \\
\text{or} \quad \lambda_0(t) = \lambda_1(t - \delta) \quad \text{for} \ t \geq t_1
\] (4.2)

Since \( \lambda_1(t) \) is the failure rate of a lognormal distribution, it follows that the failure rate of the surviving sub-population of \( S_0 \), i.e., \( \lambda_0(t_1 + x) \), is that of a lognormal. The use of a lognormal is justified because intrinsically, electromigration is a multiplicative degradation process (see section 2.3.2). Since this intrinsic property is unaffected by a change in current density, it is reasonable to assert that the TTF statistics of the surviving population are described by a lognormal.

Thus, it is proposed that the statistics of the surviving population of \( S_0 \) be obtained by shifting the origin of the lognormal that gives rise to \( \lambda_1(t) \) by \( \delta \) so that the continuity of \( F_0(t) \) at \( t = t_1 \) is maintained, as shown in Fig. 4.1d. Hence, for \( t > t_1 \), the statistics of \( S_0 \) are described by a section of a shifted lognormal distribution, the mean of which is identical to the mean of the lognormal that gave rise to \( \lambda_1(t) \). This key point motivates the first two assumptions to be made in the next section.

4.3.2 Assumptions

We now propose a set of assumptions motivated from the arguments in the previous section. These assumptions are then used to rigorously determine the CDF of \( S_0 \) for \( t > t_1 \) in the single-step case and the CDF in the general case of multiple change in currents.

Consider a population set \( S \) that has already been subjected to some prior current density stress, i.e. it is not fresh. The following mild assumptions are made about \( S \) for some time \( t_k \neq 0 \):

**Assumption 4.1.** The statistics of the TTFs for the surviving population of \( S \) are described by a (section of) shifted lognormal distribution.
Assumption 4.2. The mean of the shifted lognormal distribution (relative to its start time) is given by Black’s equation, with $J$ being the current density at time $t_k$.

Assumption 4.3. The value of $\sigma_{\ln}$ for $S$ at time $t_k$ is the same as that of the fresh population at $t = 0$.

We note that assumption 4.3 can be dropped if the dependence of $\sigma_{\ln}$ with regard to the damage accumulated due to EM is known beforehand (see section 4.3.7).

4.3.3 Determining $F_0(t)$ and $\delta$ for single-step case

Based on the assumptions proposed in the previous section, we now rigorously obtain the CDF of $S_0$ for the single step case. Define two RVs $T_0$ and $T_1$ as follows:

1. $T_0$ describes the TTF distribution of $S_0$ when it is subjected to $J_0$ for all $t \geq 0$.

2. $T_1$ describes the TTF distribution when $S_0$ is subjected to zero current density for $t \leq \delta$ and $J_1$ for all $t > \delta$. In other words, the origin of the distribution is shifted to $t = \delta$.

Clearly, the CDFs $F_{T_0}(t)$ and $F_{T_1}(t)$ of $T_0$ and $T_1$ are known:

1. $T_0$ has a lognormal distribution as given in (2.23).

2. $T_1$ has a shifted-lognormal distribution as given in (2.24) with $\Delta = \delta$.

The MTF (relative to the origin of the distribution in each case) is given by Black’s equation and the variance is experimentally determined. Using the proposed assumptions, the CDF of $S_0$ when subjected to a current-density profile of (4.1) is:

$$F_0(t) = \begin{cases} F_{T_0}(t), & 0 \leq t \leq t_1 \\ F_{T_1}(t - \delta), & t_1 < t < \infty \end{cases}$$

(4.3)

where the time-shift $\delta \in (-\infty, t_1)$ is necessary for maintaining the continuity constraint:

$$\Phi \left[ \frac{\ln(t_1 - \delta) - \mu_{\ln,1}}{\sigma_{\ln,1} \sqrt{2}} \right] = \Phi \left[ \frac{\ln t_1 - \mu_{\ln,0}}{\sigma_{\ln,0} \sqrt{2}} \right]$$

(4.4)

where $\Phi$ the is standard normal CDF, $\mu_{\ln,k} = E[\ln T_k]$ and $\sigma_{\ln,k}^2 = \text{Var}(\ln T_k)$. The CDF $F_0(t)$ needs to be continuous because the TTF of a conductor can assume any value in a given time-range.
Define $\mu_{T,k} = E[T_k]$. Since $\Phi$ is monotonic, the equality will hold provided the terms in brackets are equal. Thus:

$$\left( \frac{\ln(t_1 - \delta) - \mu_{\ln,1}}{\sigma_{\ln,1} \sqrt{2}} \right) = \left( \frac{\ln t_1 - \mu_{\ln,0}}{\sigma_{\ln,0} \sqrt{2}} \right)$$

(4.5)

Use (2.26) in (4.5) to arrive at:

$$\ln(t_1 - \delta) - (\ln \mu_{T,1} - 0.5\sigma_{\ln,1}^2) = \ln t_1 - (\ln \mu_{T,0} - 0.5\sigma_{\ln,0}^2)$$

$$\frac{1}{\sigma_{\ln,1}} \ln \left( \frac{t_1 - \delta}{\mu_{T,1}} \right) + 0.5\sigma_{\ln,1} = \frac{1}{\sigma_{\ln,0}} \ln \left( \frac{t_1}{\mu_{T,0}} \right) + 0.5\sigma_{\ln,0}$$

$$\ln \left( \frac{t_1 - \delta}{\mu_{T,1}} \right) = \frac{\sigma_{\ln,1}}{\sigma_{\ln,0}} \ln \left( \frac{t_1}{\mu_{T,0}} \right) + 0.5\sigma_{\ln,1}(\sigma_{\ln,0} - \sigma_{\ln,1})$$

(4.6)

Let $\varphi = \frac{\sigma_{\ln,1}}{\sigma_{\ln,0}}$. Then, (4.6) becomes:

$$\delta = t_1 - \mu_{T,1} \left( \frac{t_1}{\mu_{T,0}} \right)^\varphi \exp \left( \frac{\sigma_{\ln,0}^2 \varphi(1 - \varphi)}{2} \right)$$

(4.7)

Using Assumption 4.3, $\sigma_{\ln,0} = \sigma_{\ln,1}$ or $\varphi = 1$. This leads to:

$$\delta = t_1 \left[ 1 - \frac{\mu_{T,1}}{\mu_{T,0}} \right]$$

(4.8)

From Black’s equation (2.1), it is known that $\frac{\mu_{T,1}}{\mu_{T,0}} = \left( \frac{J_0}{J_1} \right)^n$, the relative time-shift $\delta$ between $T_1$ and $T_0$ is:

$$\delta = t_1 \left[ 1 - \left( \frac{J_0}{J_1} \right)^n \right]$$

(4.9)

### 4.3.4 The Case of Multiple changes in Currents

Consider a second thought experiment with $S_0$ for which the current density profile is given as:

$$J(t) = \begin{cases} 
J_0, & 0 \leq t \leq t_1 \\
J_1, & t_1 < t \leq t_2 \\
\vdots \\
J_{k-1}, & t_{k-1} < t \leq t_k \\
J_k, & t_k < t \leq t_{k+1} \\
\vdots
\end{cases}$$

(4.10)
where $J_{k-1} \neq J_k \forall k > 0$. It is interesting to note that (4.10) is the typical current density profile of a surviving interconnect in the power grid, where the $k^{th}$ failing interconnect has $\tau = t_k$. The analysis in this section assumes $J_k L > \beta_c \forall k$, i.e all the interconnects are EM-susceptible for all time-spans. The Blech-effect is incorporated in the next section.

The statistics of $S_0$ are derived based on the stated assumptions in the previous section. As per Assumption 4.1, for each time-span $t_k < t \leq t_{k+1}$, the statistics are described by a RV $T_k$ that has a lognormal distribution originating at some $t = \Delta_k$. Assumption 4.2 dictates that the mean $\mu_{T,k} = E[T_k]$ is given by Black’s Eq. with $J = J_k$. $\Delta_k$ is determined as follows: in order to satisfy the continuity constraint, each RV $T_k$ has a finite time-shift of $\delta_k$ with respect to $T_{k-1}$, with $T_0$ having a shift of $\delta_0 = 0$. This clearly implies that if the distribution for $T_k$
originates at time $\Delta_k$ with respect to $t = 0$, then $\Delta_k$ is:

$$\Delta_k = \sum_{i=0}^{k} \delta_i$$

The time-shift $\delta_k$ between $T_k$ and $T_{k-1}$ can be found using the continuity constraint:

$$\Phi\left[\ln\left(t_k - \Delta_k - \mu_{\ln,k}\right) - \frac{\mu_{\ln,k} \sqrt{2}}{\sigma_{\ln,k}}\right] = \Phi\left[\ln\left(t_k - \Delta_{k-1} - \mu_{\ln,k-1}\right) - \frac{\mu_{\ln,k-1} \sqrt{2}}{\sigma_{\ln,k-1}}\right]$$

Using $\Delta_k = \Delta_{k-1} + \delta_k$, assumption 4.3 ($\sigma_{\ln,0} = \ldots = \sigma_{\ln,n} = \sigma_{\ln}$) and (2.26) as before for the single step case, we have:

$$\ln\left(\frac{t_k - \Delta_{k-1} - \delta_k}{\mu_{T,k}}\right) = \ln\left(\frac{t_k - \Delta_{k-1}}{\mu_{T,k-1}}\right)$$

By equating the terms in brackets, this can be simplified to:

$$\delta_k = (t_k - \Delta_{k-1})(1 - r_k) = \left(t_k - \sum_{i=1}^{k-1} \delta_i\right)(1 - r_k) \quad (4.11)$$

where $r_k = \left(\frac{\mu_{T,k}}{\mu_{T,k-1}}\right) = \left(\frac{J_{k-1}}{J_k}\right)^n$. The cdf of $S_0$ can now be written as:

$$F_0(t) = \begin{cases} 
F_{T_0}(t - \Delta_0), & 0 \leq t \leq t_1 \\
F_{T_1}(t - \Delta_1), & t_1 < t \leq t_2 \\
\vdots & \\
F_{T_{k-1}}(t - \Delta_{k-1}), & t_{k-1} < t \leq t_k \\
F_{T_k}(t - \Delta_k), & t_k < t \leq t_{k+1} \\
\vdots & 
\end{cases} \quad (4.12)$$

Figure 4.2 shows a typical CDF plot of $S_0$ for multi-step current density profile as per the proposed approach. From the figure, it can be easily observed that

- if $J_k > J_{k-1}$, the CDF slope increases to mirror the fact that the rate of degradation has increased after an increase in the current density, and

- if $J_k < J_{k-1}$, the CDF slope decreases to mirror the fact that the rate of degradation has decreased after a decrease in the current density.

Note that the proposed assumptions are robust, in the sense that if $J_{k-1} = J_k$, $\delta_k = 0$ and the CDF of $S_0$ is given by a single lognormal distribution with MTF as obtained from Black’s
Chapter 4. Estimating EM statistics for step currents

The previous analysis assumed $J_k L > \beta_c \forall k$, where $\beta_c$ is given by (see (2.5)):

$$\beta_c = \frac{\Omega}{Z^* e \rho} (\Delta \Lambda)$$  \hspace{1cm} (4.13)

$\Omega$ is the atomic volume, $\rho$ is the resistivity of the material, $Z^*$ is the effective atomic charge, $e$ is the fundamental atomic charge, $\Delta \Lambda$ is the stress difference within the interconnect. In the next section, we examine a more general scenario and incorporate the Blech effect into our analysis.

4.3.5 Incorporating the Blech Effect

Let $M$ and $B$ be the set of integers $k$, where $k$ denotes the time-span $t_k < t \leq t_{k+1}$, so that $M = \{k : J_k L \leq \beta_c\}$ is the set of EM-immune time-spans and $B = \{k : J_k L > \beta_c\}$ is the set of
EM-susceptible time-spans. Clearly, $M \cap B = \emptyset$ and $M \cup B$ is the entire time period. In the previous section, it was assumed that $M = \emptyset$. Now consider a new thought experiment with $S_0$ where EM-immune and EM-susceptible time-spans interspersed with each other.

Consider a general scenario in which $(p - 1)$ consecutive EM-immune time-spans are sandwiched between two EM-susceptible time-spans. To be precise, $\{k - p, k\} \in B$ and $\{k - p + 1, \ldots k - 1\} \in M$. The surviving conductors (with $\tau > t_{k-p+1}$), will not fail in the time-spans $\{k - p + 1, \ldots k - 1\}$ because the current densities $J_{k-p+1}, \ldots J_{k-1}$ cannot generate sufficient stress for failure, regardless of state of degradation. Hence, the reliability/CDF of the population $S_0$ in the EM-immune time-spans does not change and remains the same as it was at the on-set of the corresponding EM-immune time-span. Based on this observation, the framework developed in the last section is extended to incorporate the Blech effect by introducing two modifications, as detailed below.

First, assumption 4.1 is now applicable only for EM-susceptible time-spans, so that RV $T_k$, that has a shifted lognormal distribution originating at $t = \Delta_k$, exists for $k \in B$. Since the surviving conductors do not fail for $k \in M$, the corresponding probability of failure is zero and the associated CDF is a constant function. As $k \in B$ is encountered, the conductors start failing again. Thus, the CDF of $S_0$ in this case is defined to be:

$$F_0(t) = \begin{cases} 0 & k \in M : k < \min(B) \\ F_{T_b}(t_{b+1} - \Delta_b), & k \in M, b \in B : b < k \text{ and } |k - b| \text{ is minimum} \\ F_{T_k}(t - \Delta_k), & k \in B \end{cases} \quad (4.14)$$

Second, the time of origin of the distribution for $T_k$ is modified to take into account the EM-immune time-spans:

$$\Delta_k = \Theta(t_k) + \sum_{i=0, i \in B}^{k} \delta_i \quad (4.15)$$

where $\Theta(t)$ is the sum of all EM-immune time-spans up to the present time $t$, i.e.:

$$\Theta(t) = \sum_{k \in Q} (t_{k+1} - t_k) \text{ s.t. } Q = \{k : k \in M \text{ and } t_{k+1} < t\}$$

and $\delta_k$ is time-shift of $T_k$ relative to $T_{k-p}$ needed to maintain the continuity constraint if the in-between EM-immune time-spans $\{k - p + 1, \ldots k - 1\}$ are removed. $\delta_k$ is found using the continuity constraint, which in this case is:

$$\Phi \left[ \frac{\ln(t_k - \Delta_k) - \mu_{\ln,k}}{\sigma_{\ln,k} \sqrt{2}} \right] = \Phi \left[ \frac{\ln(t_{k-p+1} - \Delta_{k-p}) - \mu_{\ln,k-p}}{\sigma_{\ln,k-p} \sqrt{2}} \right] \quad (4.16)$$

Equating the terms in the brackets and using assumption 4.3 ($\sigma_{\ln,0} = \ldots = \sigma_{\ln,n} = \sigma_{\ln}$) and
(2.26) as before, we get:

\[
\ln \left( \frac{t_k - \Delta_k}{\mu_{T,k}} \right) = \ln \left( \frac{t_{k-p+1} - \Delta_{k-p}}{\mu_{T,k-p}} \right) \tag{4.17}
\]

Since logarithm is also a monotonic function, we can equate the terms in brackets to get:

\[
t_k - \Delta_k = (t_{k-p+1} - \Delta_{k-p}) r_k \tag{4.18}
\]

where \( r_k = \left( \frac{\mu_{T,k}}{\mu_{T,k-p}} \right) = \left( \frac{J_{k-p}}{J_k} \right)^n \) in this case. From definition, it is known that:

\[
\Delta_{k-p} = \Theta(t_{k-p}) + \sum_{i=0, i \in B}^{k-p} \delta_i \tag{4.19}
\]

Subtracting (4.19) from (4.15), \( \Delta_k \) can be expressed in terms of \( \Delta_{k-p} \):

\[
\Delta_k - \Delta_{k-p} = \Theta_k - \Theta_{k-p} + \sum_{i=0, i \in B}^{k} \delta_i - \sum_{i=0, i \in B}^{k-p} \delta_i \\
= (t_k - t_{k-p+1}) + \delta_k \\
\text{or} \\
\Delta_k = \Delta_{k-p} + (t_k - t_{k-p+1}) + \delta_k \tag{4.20}
\]

Substituting the value of \( \Delta_k \) in (4.18):

\[
t_k - (\Delta_{k-p} + (t_k - t_{k-p+1}) + \delta_k) = (t_{k-p+1} - \Delta_{k-p}) r_k \\
(t_{k-p+1} - \Delta_{k-p}) - \delta_k = (t_{k-p+1} - \Delta_{k-p}) r_k \\
\delta_k = (t_{k-p+1} - \Delta_{k-p})(1 - r_k) \tag{4.21}
\]

Clearly, (4.21) reduces to (4.11) for \( p = 1 \). Figure 4.3 depicts the proposed model.

### 4.3.6 Updating A TTF sample

Consider a conductor \( C \) of the set \( S_0 \) subjected to the current density profile (4.10). Clearly, the TTF of \( C \) changes because the RV describing the statistics of the population changes for each time-span \( t_k < t \leq t_{k+1} \). At \( t_0(=0) \), \( C \) has a TTF given by (using (2.30)):

\[
\tau_0 = \mu_{T,0} \cdot \exp \left( \Psi \sigma_{\ln} - 0.5 \sigma_{\ln}^2 \right) \tag{4.22}
\]

where \( \Psi \) is a sample value from Standard Normal Distribution \( \Phi \), \( \mu_{T,0} = E[T_0] \) and \( \sigma_{\ln}^2 = Var(T_0) \). \( \sigma_{\ln} \) is constant throughout the life of conductor.
To probe further, (4.22) can be re-written as:

\[ \tau_0 = \mu_{T,0} \exp \left( \Phi^{-1}(\zeta)\sigma_{\ln} - 0.5\sigma_{\ln}^2 \right) \]  

(4.23)

where \( \zeta \) is a scalar between 0 and 1. From (4.23), one can view the process of assigning a TTF sample from a lognormal distribution as the inverse mapping of a uniformly selected random number from 0 to 1. Graphically, this is shown in figure 4.4a, where a selected point on the \( y \)-axis (\( \zeta \), a CDF value) is inversely mapped from \( F_{T_0}(t) \) to it’s corresponding value, \( \tau_0 \), on the \( x \)-axis (time axis).

When the current density changes through \( C \), the CDF is changed accordingly, as detailed in the previous text. The new TTF, \( \tau_1 \), is found in a similar way as before: \( \zeta \) is inversely mapped from \( F_{T_1}(t - \Delta_1) \) to it’s corresponding value, \( \tau_1 \), on the \( x \)-axis. This is shown in figure 4.4b. The analytical expression for the same is:

\[ \tau_1 = \Delta_1 + \mu_{T,1} \exp \left( \Phi^{-1}(\zeta)\sigma_{\ln} - 0.5\sigma_{\ln}^2 \right) \]  

(4.24)

Equation (4.24) can be generalized. Assume, for the sake of argument, that \( C \) survives for \( t > t_k \) and \( k \in B \). At \( t = t_k \), when the \( k^{th} \) current change occurs, the TTF of \( C \) is updated using the following relation:

\[ \tau_k = \Delta_k + \mu_{T,k} \exp \left( \Phi^{-1}(\zeta)\sigma_{\ln} - 0.5\sigma_{\ln}^2 \right) \]  

(4.25)

where \( \zeta \) is the same as initially used in (4.23). The offset \( \Delta_k \) (obtained from (4.20) or (4.15)) is added so that \( \tau_k \) is now referred from \( t = 0 \). For \( k \in M \), \( \tau_k \) is defined to be \( \infty \).
Finally, to verify the integrity of the proposed approach, it is proved that if the given framework is implemented, the new TTF will always be greater than the present time.

**Theorem 3.** Consider a conductor having the current density profile of (4.10). Let \( \{k-p, k\} \in B \) and \( \{k-p+1, \ldots, k-1\} \in M \). Then, if eq. (4.21) and (4.25) are used to find the offset \( (\delta_k) \) and TTF \( (\tau_k) \) for the conductor, we always have:

\[
\tau_k = t_k + (\tau_{k-p} - t_{k-p+1}) \left( \frac{J_{k-p}}{J_k} \right)^n
\]

so that \( \tau_k > t_k \).

**Proof.** Let \( \Upsilon = \exp(\Psi \sigma_{\ln} - 0.5\sigma_{\ln}^2) \). By Assumption 4.3, \( \Upsilon \) will be the same throughout the life-time of the conductor.

As stated, the conductor was EM-susceptible at \( t = t_{k-p} \), became EM-immune for all time span(s) following \( t = t_{k-p+1} \) until \( t = t_k \), when it becomes EM-susceptible again. Clearly, \( \tau_{k-p} > t_{k-p+1} \), or the conductor should have failed before \( t_{k-p+1} \). From theory, the RVs \( T_{k-p+1}, \ldots, T_{k-1} \) are undefined. From (4.15), the time of origin of \( T_k \) can be written as:

\[
\Delta_k = \Delta_{k-p} + \Theta(t_k) - \Theta(t_{k-p}) + \delta_k
\]

\[= \Delta_{k-p} + (t_k - t_{k-p+1}) + (t_{k-p+1} - \Delta_{k-p})(1 - r_k)\]

\[= t_k - r_k t_{k-p+1} + r_k \Delta_{k-p}\]

(i)

where \( r_k = \left( \frac{J_{k-p}}{J_k} \right)^n = \left( \frac{\mu_{T,k}}{\mu_{T,k-p}} \right) \). The new TTF at \( t = t_k \) can now be written as:

\[
\tau_k = \Delta_k + \mu_{T,k} \Upsilon
\]

\[= t_k - r_k t_{k-p+1} + r_k \Delta_{k-p} + r_k \mu_{T,k-p} \Upsilon\]  

(from (i))

\[= t_k + (\Delta_{k-p} + \mu_{T,k-p} \Upsilon - t_{k-p+1})r_k\]

\[= t_k + (\tau_{k-p} - t_{k-p+1})r_k\]  

(using (4.25))

\[\square\]

The result, that we update the residual life of a conductor after the current density changes, is a direct outcome of the assumptions 4.1 and 4.2. It makes sense because a change in current density at some time \( t_k \neq 0 \) can only affect the residual life of a conductor, the past statistics up to \( t = t_k \) are unchanged.

### 4.3.7 Dropping Assumption 4.3

A small note on the applicability of the above framework if \( \sigma_{\ln} \) is not considered constant: suppose the dependence of \( \sigma_{\ln} \) with regard to the damage accumulated due to EM is known
beforehand, so that one can determine $\sigma_{ln}$ based on some measure of the state of degradation of the surviving conductors at the time of the current-density change. The new shifted lognormal distribution (used to describe the statistics of the surviving population after the current-density change) will thus have the new standard deviation. In such a case, we can argue that since CDF of a lognormal is a monotonically increasing function in time, the updated TTF determined using the first two assumptions and the continuity constraint will always be greater than the present time. For example, in figure 4.4b, $\tau_1 > t_1$ even if the CDF $F_{T_1}(t - \Delta_1)$ has a different standard deviation.

4.4 Implementation

4.4.1 Obtaining a grid TTF sample

The overall flow for obtaining a sample of power grid TTF using the series and mesh model is given in Algorithm 2. The inputs to the algorithm are the factorized upper and lower triangular matrices of $G_0$ ($L_G, U_G$), the list of all interconnects in the grid ($R_{list}$), the vector of initial voltage drops at $t = 0$ ($v_0$), the voltage drop threshold vector for all nodes of the grid ($v_{th}$), the critical Blech product ($\beta_c$) and a flag `enable_ttf_update` which controls if the TTF updates are to be performed while obtaining the grid TTF sample. The outputs of the algorithm are $\tau_s$ and $\tau_m$, the grid TTF sample considering it as a series and mesh system, respectively.

Using a random number generator, sample TTFs are assigned to all the interconnects in the power grid based on their current densities at $t = 0$. Considering the grid as a series system, failure of the first resistor should cause the grid to fail. Hence $\tau_s$ is assigned the TTF sample of the first resistor in the sorted list $R_{list}$. In order to find the TTF sample of the grid using the mesh model ($\tau_m$), we start failing resistors. The voltage drops are efficiently updated using the Woodbury-Banachiewicz formulation (Algorithm 1) and the TTFs of surviving interconnects are updated as outlined in section 4.3.5. The updated TTFs may change the sequence of failures, hence the surviving interconnects are re-ordered as per their updated TTFs to determine the next interconnect to fail. The interconnects keep failing until either node voltage drop(s) exceeds $v_{th}$ or the grid becomes singular, after which the grid is deemed to have failed. $\tau_m$ is assigned the TTF sample of the last resistor, which caused the grid to fail.

4.4.2 Estimating grid MTF

The MTF of a grid, for both the series and the mesh model, is estimated by random sampling. In other words, Algorithm 2 is run $w$ times to generate $w$ grid TTF samples. The arithmetic mean of these samples is the estimate of the true mean $\mu$. The number of samples $w$ is decided based on the stopping criteria derived in section 2.5.1. The parameter values chosen in this case are $\alpha = 0.05$ and $\epsilon = 0.05$, i.e. there is a 95% confidence that the estimated MTF is within
Algorithm 2 TTF\_GRID

Input: $L_G, U_G, R_{\text{list}}, v_0, v_{th}, \beta_c, \text{enable}_\text{ttf}_\text{update}$

Output: $\tau_m, \tau_s$ \hspace{1cm} \triangleright \text{sample grid TTF using mesh and series model}

1: Use Random Number Generator to assign TTF samples to all resistors in the list $R_{\text{list}}$ based on current densities at $t = 0$, as per (4.22).
2: Sort $R_{\text{list}}$ in ascending order of TTF samples
3: $\tau_s \leftarrow R_{\text{list}}[1].\text{ttf}$
4: for $k = 1 \rightarrow \text{SIZE}(R_{\text{list}})$ do
5: Find $\Delta G_k$, the conductance stamp of $R_{\text{list}}[k]$, as per (3.3).
6: \{$v_k, \text{grid}_\text{singular}$\} $\leftarrow \text{WB}(v_0, \Delta G_k, L_G, U_G, Z, H, W_{\text{inv}}, y, \gamma, k)$
7: if $\text{grid}_\text{singular} = 1$ then
8: $\tau_m \leftarrow R_{\text{list}}[k].\text{ttf}$ \hspace{1cm} \triangleright \text{grid failed due to singularity}
9: return \hspace{1cm} \triangleright \text{return to callee}
10: end if
11: for $q = 1 \rightarrow \text{SIZE}(v_k)$ do
12: if $v_k[q] > v_{th}[q]$ then
13: $\tau_m \leftarrow R_{\text{list}}[k].\text{ttf}$ \hspace{1cm} \triangleright \text{grid failed because } v(t) > v_{th}$
14: return \hspace{1cm} \triangleright \text{return to callee}
15: end if
16: end for
17: if $\text{enable}_\text{ttf}_\text{update} = 1$ then
18: for $q = k + 1 \rightarrow \text{SIZE}(R_{\text{list}})$ do
19: Update the TTF of $R_{\text{list}}[q]$ as outlined in section 4.3.5.
20: end for
21: Sort $R_{\text{list}}$ from $k + 1 \rightarrow \text{SIZE}(R_{\text{list}})$ in ascending order of TTF samples
22: end if
23: end for

5% of the true MTF for the respective model. Algorithm 3, MTF\_GRID, computes the MTF as per the series and mesh models.

4.5 Experimental Results

A \texttt{C++} implementation was written based on the proposed approach to estimate grid MTF as per the series and mesh models. As input, two types of test grids were used. The first type were \textit{generated} per user specifications, including grid dimensions, metal layers, pitch and width per layer. The supply voltages and current sources were randomly placed on the grid. The technology specifications were consistent with 1.1 V 65nm CMOS technology. These grids are henceforth referred to as \textit{internal} grids. The second type of grids are part of IBM power grid benchmarks [52]. These grids are \textit{dual grids} (i.e. they have both $v_{dd}$ and $gnd$ rails of the power grid), but the proposed approach was tested only for $v_{dd}$ part of the grids and are referred to as \textit{external} grids. The grid details are given in Table 4.1.

The \textit{Voltage drop threshold} was defined to be 10% of $v_{dd}$ for all nodes in a grid [44]. For
Algorithm 3 \texttt{MTF\_GRID}

\textbf{Input:} $G_0, i, R_{\text{list}}, v_{th}, \beta_c$, enable\_ttf\_update

\textbf{Output:} $\mu_m, \mu_s$ \hspace{1cm} \text{\scriptsize{\textgreater} estimated grid MTF using mesh and series model}

1: $\{L_G, U_G\} \leftarrow \text{SPARSE\_LU}(G_0)$
2: $v_0 \leftarrow \text{BF\_SUBSTITUTION}(L_G, U_G, i)$ \hspace{1cm} \text{\scriptsize{\textgreater} backward-forward substitution}
3: for $p = 1 \rightarrow \text{SIZE}(R_{\text{list}})$ do
4: \hspace{1cm} Calculate the initial current density $J_{0,p}$ through resistor $R_{\text{list}}[p]$
5: \hspace{1cm} Using $\beta_c$, calculate and store $J_{c,p}$, the critical current density for $R_{\text{list}}[p]$ below which EM failure will not occur
6: \hspace{1cm} if $J_{0,p} < J_{c,p}$ then
7: \hspace{1cm} \hspace{1cm} $R_{\text{list}}[p].\mu_{T,0}$ $\leftarrow \infty$
8: \hspace{1cm} \hspace{1cm} else
9: \hspace{1cm} \hspace{1cm} Calculate and store $R_{\text{list}}[p].\mu_{T,0}$ based on Black’s Equation (2.1)
10: \hspace{1cm} end if
11: \hspace{1cm} end for
12: $p \leftarrow 0$ \hspace{1cm} \text{\scriptsize{\textgreater} initialize iteration counter}
13: $Z \leftarrow 1.96$, $\epsilon \leftarrow 0.05$ \hspace{1cm} \text{\scriptsize{\textgreater} parameters to calculate stopping criteria}
14: $w_s \leftarrow 50$, $w_m \leftarrow 50$
15: $\mu_s \leftarrow 0$, $\sigma_s \leftarrow 1$, $\mu_m \leftarrow 0$, $\sigma_m \leftarrow 1$
16: while $p \leq \text{MAX}(w_s, w_m)$ do
17: \hspace{1cm} Reset the power grid to its initial state
18: \hspace{1cm} $p \leftarrow p + 1$ \hspace{1cm} \text{\scriptsize{\textgreater} increment iteration counter}
19: \hspace{1cm} $\{\tau_m[p], \tau_s[p]\} \leftarrow \text{TTF\_GRID}(G_0, L_G, U_G, R_{\text{list}}, v_0, v_{th}, \beta_c, \text{enable\_ttf\_update})$
20: \hspace{1cm} $\mu_s \leftarrow \text{MEAN}(\tau_s)$, $\sigma_s \leftarrow \text{STD\_DEV}(\tau_s)$
21: \hspace{1cm} $\mu_m \leftarrow \text{MEAN}(\tau_m)$, $\sigma_m \leftarrow \text{STD\_DEV}(\tau_m)$
22: \hspace{1cm} $w_s \leftarrow \text{MAX}\left(\left(\frac{Z\sigma_s}{\mu_s\epsilon/(1+\epsilon)}\right)^2, 50\right)$ \hspace{1cm} \text{\scriptsize{\textgreater} stopping criteria for series model}
23: \hspace{1cm} $w_m \leftarrow \text{MAX}\left(\left(\frac{Z\sigma_m}{\mu_m\epsilon/(1+\epsilon)}\right)^2, 50\right)$ \hspace{1cm} \text{\scriptsize{\textgreater} stopping criteria for mesh model}
24: end while

Concreteness in results, the following configuration was assumed: the interconnect material is Aluminum (Al), with activation energy $E_a = 0.9eV$, $\beta_c = 3000A/cm$ [53] and a current exponent $n = 1$, assuming a growth dominated failure due to the presence of shunt layers [54]. The standard deviation of the lognormal $\sigma_{ln}$ was assumed to be 0.81 for all interconnects in the grid, consistent with typical data in the literature [30] [31]. Since we use an empirical model to estimate the interconnect MTF, temperature enters into the analysis through Black’s equation. In this work, a nominal interconnect temperature of 373K was used in all simulations. Note that the proposed implementation can incorporate temperature/thermal profiles (obtained by using [55] or some other approach) as well. After each interconnect failure, the current density changes and hence the thermal profile needs to be updated.
A 2.6 GHz Linux machine with 24 GB of RAM was used for all simulations. The experiments were carried out with two objectives: a) To establish that the mesh model reduces the pessimism in MTF estimation of the power grid, and b) To bring out the importance of the TTF update (TTFU) framework proposed in this chapter. To accomplish the same, algorithm $\text{MTF}_\text{GRID}$ was run twice on each test grid, first with $\text{enable}_\text{ttf updates} = 1$ and second time with $\text{enable}_\text{ttf updates} = 0$. Table 4.2 tabulates the results of the first run, in which the statistics of all surviving interconnects are updated after each failure. This is the original approach, and the results from this run serve as a de-facto standard of comparison for other derivative approaches. Table 4.3 shows the results of the second run, in which the statistics are not updated for any interconnect failure. This is the secondary approach. Henceforth, $\mu_m^*$ denotes the grid MTF estimated from the original approach, and $\mu'_m$ denotes the grid MTF estimated from the secondary approach.

The total time taken by either approach can be empirically expressed as:

$$t_{\text{cpu}} = t_{\text{init}} + \sum_{p=1}^{w_m} t_{mc,p} + t_{\text{extra}}$$

where $t_{\text{cpu}}$ is the total time, $t_{\text{init}}$ is the initial one-time processing done to obtain the LU factorization and store the initial branch MTFs, $t_{mc,p}$ is the time taken to complete the $p^{th}$ Monte-carlo iteration and $t_{\text{extra}}$ is the extra time spent in obtaining grid TTF samples as per
the series model if $w_s > w_m$. In practice, $t_{\text{init}}$ and $t_{\text{extra}}$ are very small as compared to the second term, hence

$$t_{\text{cpu}} \approx \sum_{p=1}^{w_m} t_{mc,p} = w_m \times \bar{t}_{mc} = w_m \times \bar{t}_f \times \bar{k}_f$$

(4.27)

where $\bar{t}_{mc}$ is the average time per Monte-carlo iteration, $\bar{t}_f$ is the average time spent in updating voltage drops and TTFs after an interconnect fails and $\bar{k}_f$ is the average number of interconnect failures required to fail the power grid.

**Comparison between series and mesh model** The power grid MTF as estimated using the series model and mesh model is compared using a gain ratio ($\mu^*_{m}/\mu_{s}$). Among other things, the gain ratio is dependent on $\max(v_0)$, the maximum node voltage drop in the grid at $t = 0$. If the difference between $\max(v_0)$ and $v_{th}$ increases, the gain ratio also increases. If the difference is small, then the mesh model degenerates to series model. Given a reasonable difference between $v_{th}$ and $\max(v_0)$, the gain ratio is found to be 3-4 for all the test grids, the average gain ratio being 3.78 for the test grids. Figure 4.5 graphically compares $\mu_s$ and $\mu^*_{m}$ for the test grids using a horizontal bar graph. It is observed that on average, large grids can tolerate upto 115 interconnect failures before the grid truly fails. These results clearly indicate that the series model, that declares a grid to be failed just after the first interconnect failure, is highly pessimistic for EM reliability verification.

**Significance of TTF update framework** Next, for the secondary run, the TTF updates are turned off. The accuracy of results and performance of the estimation engine is compared to the original approach, and is reported in table 4.3. The relative speed-up is expressed as a ratio of time taken in the original approach to the time taken in the secondary approach. The error in estimation is measured as

$$\varepsilon' \triangleq \frac{(\mu^*_{m} - \mu^*_{m}) \times 100}{\mu^*_{m}} \%$$

The estimated MTF $\mu^*_{m}$ in the secondary approach is optimistic for all the test grids as compared to original approach. The average error in estimation is found to be 19.6%. Intuitively, $\varepsilon'$ should be proportional to both $v_{\text{margin}} = v_{th} - \max(v_0)$ and $\sigma_m$, because $v_{\text{margin}}$ and $\sigma_m$ control the length and the re-ordering in the sequence of interconnect failures because of TTF updates in a given realization. If these parameters are small, the error should be small. Figure 4.6 clearly confirms this intuition.
Table 4.2: Comparison of Power Grid MTF $\mu_s$ and $\mu_m^*$ as obtained using the *Series* model and *Mesh* model for the original approach. $\mu_s$ and $\mu_m^*$ are estimated with 95% confidence and maximum 5% relative error.

<table>
<thead>
<tr>
<th>Grid Name</th>
<th>max($v_0$) (% $v_{dd}$)</th>
<th>Series Model</th>
<th>Mesh Model with TTF updates turned on</th>
<th>Gain ($\mu_m^*/\mu_s$)</th>
<th>CPU Time $t_{cpu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_s$ (yrs)</td>
<td>$#MC$</td>
<td>$\mu_m^*$ (yrs)</td>
<td>$#MC$</td>
<td>$#Fails$</td>
</tr>
<tr>
<td>IBMPG2</td>
<td>7.0%</td>
<td>10.18</td>
<td>183</td>
<td>34.06</td>
<td>50</td>
</tr>
<tr>
<td>IBMPG3</td>
<td>6.8%</td>
<td>9.48</td>
<td>161</td>
<td>40.01</td>
<td>86</td>
</tr>
<tr>
<td>IBMPG4</td>
<td>7.4%</td>
<td>10.28</td>
<td>122</td>
<td>37.65</td>
<td>51</td>
</tr>
<tr>
<td>IBMPG5</td>
<td>4.8%</td>
<td>9.96</td>
<td>193</td>
<td>32.34</td>
<td>89</td>
</tr>
<tr>
<td>IBMPG6</td>
<td>5.7%</td>
<td>10.47</td>
<td>171</td>
<td>41.93</td>
<td>76</td>
</tr>
<tr>
<td>IBMPGNEW1</td>
<td>9.4%</td>
<td>12.94</td>
<td>172</td>
<td>59.96</td>
<td>92</td>
</tr>
<tr>
<td>IBMPGNEW2</td>
<td>7.7%</td>
<td>10.12</td>
<td>164</td>
<td>39.64</td>
<td>75</td>
</tr>
<tr>
<td>G1</td>
<td>3.6%</td>
<td>10.12</td>
<td>186</td>
<td>33.64</td>
<td>87</td>
</tr>
<tr>
<td>G2</td>
<td>4.6%</td>
<td>10.27</td>
<td>132</td>
<td>33.7</td>
<td>76</td>
</tr>
<tr>
<td>G3</td>
<td>5.0%</td>
<td>10.89</td>
<td>115</td>
<td>37.87</td>
<td>82</td>
</tr>
<tr>
<td>G4</td>
<td>4.6%</td>
<td>12.03</td>
<td>121</td>
<td>49.92</td>
<td>56</td>
</tr>
<tr>
<td>G5</td>
<td>4.0%</td>
<td>10.48</td>
<td>86</td>
<td>42.90</td>
<td>50</td>
</tr>
</tbody>
</table>

Average Gain ratio: 3.78
Table 4.3: Speed and accuracy of the secondary approach, $\mu'_m$ is estimated with 95% confidence and maximum 5% relative error.

<table>
<thead>
<tr>
<th>Grid Name</th>
<th>max($v_0$)</th>
<th>$\mu'_m$ (yrs)</th>
<th>$w_m$</th>
<th>$k_f$</th>
<th>$t_f$</th>
<th>$t_{mc}$</th>
<th>$t_{cpu}$</th>
<th>$\varepsilon'$</th>
<th>Speed-up in $t_f$</th>
<th>Speed-up in $t_{cpu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBMPG2</td>
<td>7.0%</td>
<td>35.72</td>
<td>64</td>
<td>19.50</td>
<td>0.15s</td>
<td>3.01s</td>
<td>3.21m</td>
<td>↑ 4.88%</td>
<td>1.91x</td>
<td>1.54x</td>
</tr>
<tr>
<td>IBMPG3</td>
<td>6.8%</td>
<td>44.97</td>
<td>89</td>
<td>51.73</td>
<td>1.51s</td>
<td>1.30m</td>
<td>1.93h</td>
<td>↑ 12.39%</td>
<td>2.01x</td>
<td>1.75x</td>
</tr>
<tr>
<td>IBMPG4</td>
<td>7.4%</td>
<td>40.74</td>
<td>53</td>
<td>44.72</td>
<td>2.40s</td>
<td>1.79m</td>
<td>1.58h</td>
<td>↑ 8.21%</td>
<td>1.6x</td>
<td>1.43x</td>
</tr>
<tr>
<td>IBMPG5</td>
<td>4.8%</td>
<td>34.93</td>
<td>141</td>
<td>18.05</td>
<td>0.61s</td>
<td>10.94s</td>
<td>25.72m</td>
<td>↑ 8.00%</td>
<td>2.63x</td>
<td>1.59x</td>
</tr>
<tr>
<td>IBMPG6</td>
<td>5.7%</td>
<td>48.96</td>
<td>96</td>
<td>35.92</td>
<td>1.11s</td>
<td>39.78s</td>
<td>63.65m</td>
<td>↑ 16.77%</td>
<td>2.97x</td>
<td>2.14x</td>
</tr>
<tr>
<td>IBMPGNEW1</td>
<td>9.4%</td>
<td>60.97</td>
<td>93</td>
<td>64.53</td>
<td>1.20s</td>
<td>1.30m</td>
<td>2.01h</td>
<td>↑ 1.69%</td>
<td>2.29x</td>
<td>2.14x</td>
</tr>
<tr>
<td>IBMPGNEW2</td>
<td>7.7%</td>
<td>44.00</td>
<td>103</td>
<td>47.41</td>
<td>3.80s</td>
<td>3.00m</td>
<td>5.15h</td>
<td>↑ 11.00%</td>
<td>1.41x</td>
<td>0.91x</td>
</tr>
<tr>
<td>G1</td>
<td>3.6%</td>
<td>50.13</td>
<td>154</td>
<td>53.31</td>
<td>0.11s</td>
<td>6.05s</td>
<td>15.54m</td>
<td>↑ 49.04%</td>
<td>1.85x</td>
<td>0.47x</td>
</tr>
<tr>
<td>G2</td>
<td>4.6%</td>
<td>46.16</td>
<td>110</td>
<td>71.09</td>
<td>0.32s</td>
<td>22.48s</td>
<td>41.21m</td>
<td>↑ 36.97%</td>
<td>1.75x</td>
<td>0.59x</td>
</tr>
<tr>
<td>G3</td>
<td>5.0%</td>
<td>44.86</td>
<td>106</td>
<td>78.99</td>
<td>0.62s</td>
<td>48.72s</td>
<td>1.43h</td>
<td>↑ 18.45%</td>
<td>1.78x</td>
<td>0.89x</td>
</tr>
<tr>
<td>G4</td>
<td>4.6%</td>
<td>64.10</td>
<td>94</td>
<td>185.6</td>
<td>1.75s</td>
<td>5.40m</td>
<td>8.46h</td>
<td>↑ 28.41%</td>
<td>1.73x</td>
<td>0.54x</td>
</tr>
<tr>
<td>G5</td>
<td>4.0%</td>
<td>59.67</td>
<td>59</td>
<td>269.9</td>
<td>3.64s</td>
<td>16.39m</td>
<td>16.12h</td>
<td>↑ 39.08%</td>
<td>1.61x</td>
<td>0.58x</td>
</tr>
</tbody>
</table>

Average Relative error and speed-up | 19.57% | 1.96x | 1.21x
Figure 4.5: Comparison of MTF as obtained using the series model and mesh model for the original approach

When TTF updates are turned off, no computation power is spent to update the statistics of the surviving interconnects. Figure 4.7 shows the average division of CPU time between voltage updates and TTF updates for each grid after an interconnect failure as observed from the original approach. Hence, by eliminating the TTF updates in secondary approach, there is an average speed-up of 1.96x in $t_f$. However, eliminating TTF updates slows down the convergence of Monte Carlo iterations and increases number of interconnect failures required to fail a grid. In effect, this increases $w_m$ and $k_f$. From (4.27), we have (the subscript org stands for the original approach and sec stands for the secondary approach):

$$\frac{t_{cpu,org}}{t_{cpu,sec}} = \frac{w_{m,org} \times t_{f,org} \times k_{f,org}}{w_{m,sec} \times t_{f,sec} \times k_{f,sec}}$$

Thus, for the secondary approach, if the speed-up obtained in $t_f$ is offset by slow Monte Carlo convergence and longer failure sequences, the net CPU time $t_{cpu}$ required to estimate grid MTF increases, otherwise it decreases. For the test grids, an average speed-up of 1.2x was observed in $t_{cpu}$ when TTF updates are turned off, which is very less as compared to the speed-up obtained for $t_f$.

To conclude, the TTF update framework presented in this chapter is important because it removes the optimism that creeps in when the effect of failing interconnects on the surviving ones are not considered and improves the convergence. It acknowledges that the failure sequence of interconnects is dynamic, in that it can change based on the history of failures up to the present time. However, TTF updates can be turned off especially when $\max(v_0)$ is close to $v_{th}$ and $\sigma_{ln}$ is small, since $\varepsilon'$ in this case will be small and convergence is not an issue.
Figure 4.6: Dependence of relative error between $\mu'_m$ and $\mu^*_m$ on $(v_{th} - \max(v_0))$ and $\sigma_m$ for IBMPG2 grid

**Scalability** The problem size in our case is decided by the sum of number of nodes $m$ and number of interconnect branches $b$ in the power grid, because we update the node voltage drops and the statistics of the surviving branches for each failure. However, $b = O(m)$. Hence, we use $m$ as the problem size in scalability analysis. We determine the scalability of the following for the original approach a) Average time taken to update voltage drops and statistics ($t_f$), b) Average time per Monte Carlo iteration ($t_{mc}$), c) Total time taken to estimate MTF ($t_{cpu}$) d) Maximum memory used. As shown in figure 4.8, scalability of $t_f$, $t_{mc}$ and $t_{cpu}$ are found to be $O(m^{1.19})$, $O(m^{1.61})$ and $O(m^{1.55})$, respectively. The space complexity is found to $O(m^{1.12})$.

**Voltage maps** Finally, we plot a voltage-drop map of three test grids at $t = 0$ (when they are fresh), and when they fail as per the mesh model. The maps clearly indicate the hot regions on the grid, which may be used as a graphical feedback to the user for further improving the design of the power grid.
Chapter 4. Estimating EM statistics for step currents

Figure 4.7: Average time-breakup between node voltage updates and TTF updates of surviving interconnects after an interconnect failure

Figure 4.8: Scalability of the original approach
Figure 4.9: Voltage drop maps of IBMPGNEW2 (a) at $t = 0$ and (b) at the time of failure grid failure. For this particular realization, 68 interconnect failures were required to exceed threshold voltage $v_{th} = 0.18V$

Figure 4.10: Voltage drop maps of IBMPG5 (a) at $t = 0$ and (b) at the time of failure grid failure. For this particular realization, 20 interconnect failures were required to exceed threshold voltage $v_{th} = 0.18V$
Figure 4.11: Voltage drop maps of IBMPG6 (a) at $t = 0$ and (b) at the time of failure grid failure. For this particular realization, 36 interconnect failures were required to exceed threshold voltage $v_{th} = 0.18V$.
Chapter 5

Reliability Estimation and Selective TTF updates

5.1 Introduction

In this chapter, we first develop a framework to estimate the probability of survival of a power grid for a given period of time based on Monte Carlo random sampling approach as per the series and the mesh model. This framework enables us to perform a complete statistical analysis of the power grid as per the series and the mesh model, and draw important conclusions about the statistical nature of grid failure. Next, we describe a heuristic approach to speedup the MTF estimation engine by performing selective TTF updates, i.e updating the statistics of only significantly impacted interconnects for a given interconnect failure. The results show a $\sim 2x$ speedup with negligible loss in accuracy. Finally, it is shown that the reduction in pessimism as predicted by the mesh model is stable.

5.2 Survival Probability and Reliability Estimation

Sometimes, users are interested in finding the probability of survival of a given grid for a period of $y$ years. This is called Survival Probability Estimation of the grid for a time-period of $y$ years, denoted by $SPE(y)$. In order to estimate the probability of survival of a given grid up to a period of $y$ years, Monte Carlo random sampling is used, the details of which are given in section 2.5.2. The stopping criterion (in terms of number of samples $w$ required) is defined such that there is $(1 - \alpha) \times 100\%$ confidence that the error in probability estimation is less than $\mathcal{E}$ [36]:
\[ w = \max \left[ \left( \frac{z_{\alpha/2}}{2\varepsilon} \right)^2, \left( \frac{\sqrt{63} + z_{\alpha/2}}{2\sqrt{\varepsilon}} \right)^2, \left( \frac{z_{\alpha/2}\sqrt{2\varepsilon} + 0.1 + \sqrt{(\varepsilon + 0.1)z_{\alpha/2}^2 + 3\varepsilon}}{2\varepsilon} \right)^2 \right] \] (5.1)

where \( z_{\alpha/2} \) is as defined in section 2.5.1.

SPE Algorithm  Algorithm 4 estimates the survival probability of a given grid for a time period of \( y \) years as per both the series and mesh models using the process of Bernoulli trials: for each trial (iteration), we have a success if the grid survives for \( y \) years, or else we have a failure. If \( x \) successes were observed in \( w \) trials, the survival probability is estimated to be \( \frac{x}{w} \).

Apart from \( y \), the other inputs are the conductance matrix of the fresh grid \( G_0 \), the vector of current sources \( i \), the list of all interconnects in the grid (\( R_{\text{int}} \)), the vector of initial voltage drops at \( t = 0 \) (\( v_0 \)), the voltage drop threshold vector for all nodes of the grid (\( v_{\text{th}} \)), the critical Blech product (\( \beta_c \)) and a flag \texttt{enable_ttf_update} which controls if the TTF updates are to be performed while obtaining grid TTF sample. The outputs are \( P_s = \mathcal{P}(\tau_s \geq y) \) (probability of grid surviving more than \( y \) yrs using the series system model) and \( P_m = \mathcal{P}(\tau_m \geq y) \) (probability of grid surviving more than \( y \) yrs using the mesh system model). Both \( P_s \) and \( P_m \) are estimated with a confidence of 95\% (\( \alpha = 0.05 \)) and a maximum relative error of 5\% (\( \varepsilon = 0.05 \)), which leads to a stopping criteria of \( w = 490 \) iterations, as per (5.1).

Reliability Estimation  The reliability of a power grid is essentially its survival probability at different points in time. Hence, if the statistics of TTFs of the power grid following the mesh and series model is described by the RVs \( T_m \) and \( T_s \), then their probability distribution function and reliability can be calculated empirically by obtaining \( w \) sample grid TTF values as per the respective models and calculating the survival probabilities for different values of \( y \).

It is useful in two ways:

- **Quick analysis:** If the stopping criterion is relaxed (by increasing \( \alpha \) and \( \varepsilon \)), one can quickly verify if the given grid still satisfies the target MTF \( \mu \) (by ensuring \( P_m - \varepsilon > 0.5 \) at \( y = \mu \)). This is especially useful when the power grid design goes through incremental changes, and it is costly to run a full-fledged MTF estimation engine for every change.

- **Complete Statistical Analysis:** On the other hand, towards the end of grid design, a designer can use Algorithm 4 with a strict stopping criteria to get complete information about RVs \( T_m \) and \( T_s \), in the form of their CDF or Reliability function. A complete statistical analysis for grid IBMPG5 is presented later in this chapter.
Algorithm 4 SURVIVAL_PROBABILITY_ESTIMATION

Input: $y, G_0, i, R_{\text{list}}, v_\text{th}, \beta_c, \text{enable}_\text{ttf_update}$

Output: $\mathcal{P}_m, \mathcal{P}_s$  \hspace{1cm} $\triangleright$ SPE of the grid as per the series and mesh model

1: $\{L_G, U_G\} \leftarrow \text{SPARSE\_LU}(G_0)$
2: $v_0 \leftarrow \text{BF\_SUBSTITUTION}(L_G, U_G, i)$ \hspace{1cm} $\triangleright$ backward-forward substitution
3: for $p = 1 \rightarrow \text{SIZE}(R_{\text{list}})$ do
4: Calculate the initial current density $J_{0,p}$ through resistor $R_{\text{list}}[p]$
5: Using $\beta_c$, calculate and store $J_{c,p}$, the critical current density for $R_{\text{list}}[p]$ below which EM failure will not occur
6: if $J_{0,p} < J_{c,p}$ then
7: \hspace{1cm} $R_{\text{list}}[p].\mu T,0 \leftarrow \infty$
8: else
9: \hspace{1cm} Calculate and store $R_{\text{list}}[p].\mu T,0$ based on Black’s Equation (2.1)
10: end if
11: end for
12: $k \leftarrow 0$
13: $Z \leftarrow 1.96, \hspace{0.2cm} \mathcal{E} \leftarrow 0.05$ \hspace{1cm} $\triangleright$ 95% accuracy and maximum 0.05 absolute error
14: $x_m \leftarrow 0, \hspace{0.2cm} x_s \leftarrow 0$ \hspace{1cm} $\triangleright$ initialize the counters
15: $w \leftarrow \text{MAX} \left( \left( \frac{Z}{2\mathcal{E}} \right)^2, \left( \frac{Z\sqrt{2\mathcal{E}} + 0.1 + \sqrt{(\mathcal{E} + 0.1)Z^2 + 3\mathcal{E}}}{2\mathcal{E}} \right)^2, \left( \frac{\sqrt{63 + Z}}{2\sqrt{\mathcal{E}}} \right)^2 \right)$ \hspace{0.5cm} $\triangleright$ pre-determine the stopping criteria
16: while $k \leq \lceil w \rceil$ do
17: \hspace{1cm} $\{\tau_m, \tau_s\} \leftarrow \text{TTF\_GRID}(G_0, L_G, U_G, R_{\text{list}}, v_0, v_\text{th}, \beta_c, \text{enable}_\text{ttf_update})$
18: \hspace{1cm} if $\tau_m \geq y$ then
19: \hspace{2cm} $x_m \leftarrow x_m + 1$ \hspace{1cm} $\triangleright$ increment success count for mesh model
20: \hspace{1cm} end if
21: \hspace{1cm} if $\tau_s \geq y$ then
22: \hspace{2cm} $x_s \leftarrow x_s + 1$ \hspace{1cm} $\triangleright$ increment success count for series model
23: \hspace{1cm} end if
24: \hspace{1cm} $k \leftarrow k + 1$
25: end while
26: $\mathcal{P}_m \leftarrow \frac{x_m}{w}, \hspace{0.2cm} \mathcal{P}_s \leftarrow \frac{x_s}{w}$
5.3 Selective TTF updates for speed-up

5.3.1 Motivation

In MTF estimation and survival probability estimation of a power grid as per the mesh model, the bulk of the computation effort is spent on:

1. updating the voltage drop of nodes after an interconnect failure.
2. updating the TTFs of surviving interconnects to determine the next interconnect to fail.

Figure 4.7 (in chapter 4) shows for each grid, the computation time spent after each interconnect failure in updating voltage drops and interconnect TTFs. On average, 41.3% computation time is spent in updating voltage drops and 58.7% time is spent in updating TTFs. The secondary approach, where TTFs are not updated for any interconnect, sped-up the average time it takes to determine the next interconnect to fail ($t_{f}$). But, this speed-up was not carried over to total time $t_{cpu}$ due to slower convergence. The approach presented in this section is aimed at reducing the CPU time spent in updating the TTFs in a way that minimally affect the convergence. Another advantage of this approach is that it helps trade-off accuracy for speed, if required.

5.3.2 Overview of approach

When an interconnect in the power grid fails, not all the nodes are equally impacted. This is easily verified by figure 5.1, which shows the change in node voltage drops for a randomly chosen single interconnect failure in power grids IBMPG5, IBMPG6, IBMPGNEW2 and G4. The sensitivity of voltage drops due to interconnect failure is higher in external grids because these grids have comparatively large number of current sources distributed throughout the network. In any case, it is clearly visible that for a given interconnect failure, only a subset of the nodes (and the associated links), in the immediate neighborhood are significantly impacted. This locality effect can be exploited to update the TTFs of selective interconnects and thus, speed-up the estimation engine, at the cost of some loss in accuracy of estimation.

Let the set of all nodes be represented by $\mathcal{N}$. Each node in the power grid is represented in the set by its node number. Thus $\mathcal{N}$ is the set of integers from 1 to $m$, where $m$ is the number of nodes in the power grid:

$$\mathcal{N} = \{1, 2, 3, \ldots, m - 1, m\}$$  \hspace{1cm} (5.2)

Also, let $\Gamma(\mathcal{N}_{sub})$, where $\mathcal{N}_{sub} \subseteq \mathcal{N}$, be the set of all resistors connected to nodes $k \in \mathcal{N}_{sub}$. Note that $\Gamma(\mathcal{N})$ is the set of all resistors in the power grid. Then, if $\frac{\partial v_k}{\partial R}$ is the change in the voltage drop of the $k^{th}$ node with respect to failure of interconnect $R$, we define the region of impact of an interconnect $R$, $\pi(R)$ to be:
Figure 5.1: Change in node voltage drops for a randomly chosen single interconnect failure in power grids IBMPG5, IBMPG6, IBMPGNEW2 and G4. Nodes where change in voltage drop is less than $10^{-3}$ are omitted in the figure.

\[ \pi(R) \triangleq \Gamma(\mathcal{N}_R) \text{ such that } \mathcal{N}_R = \left\{ k \in \mathcal{N} : \frac{\partial v_k}{\partial R} > \delta^* \right\} \]  

(5.3)

$\delta^*$ is a user-input that determines the boundary of the region of impact on the power grid. Intuitively, smaller value of $\delta^*$ increases $|\pi(R)|$, which improves the accuracy of estimated MTF or survival probability at the cost of reduction in speed, and vice-versa. Figure 5.2 shows how the boundary of $\pi(R)$ varies with $\delta^*$. The TTF of a resistor connected to a node $k \in \pi(R)$ is updated if $R$ fails, otherwise its TTF is not updated. The heuristic employed here aims to reduce the error in estimation by updating the TTFs of significantly impacted interconnects only, hoping that the sequence of interconnect failures for this realization is as close as possible to the actual one obtained when the TTFs of all interconnects are updated.
The selective TTF update algorithm  Algorithm 5 gives the implementation of the proposed selective updates approach. This algorithm is similar to the Algorithm 2 (in chapter 4), except in lines 18-20, where instead of updating the TTFs of all interconnects after a failure, only those interconnects that are in the set $\pi(R)$ for a given interconnect $R$ are updated when $R$ fails.

5.4 Parameter Variation

There are many factors that influence degradation of a metal line due to electromigration and thus affect the estimated MTF of the grid as per the series and mesh model. Some of these
Algorithm 5 TTF\_GRID\_SELECTIVE

**Input:** $L_G, U_G, R_{\text{list}}, v_0, v_{th}, \beta_c, \delta v^\star, \texttt{enable\_ttf\_update}$

**Output:** $\tau_m, \tau_s$  \hspace{1cm} $\triangleright$ sample grid TTF using mesh and series model

1: Use Random Number Generator to assign TTF samples to all resistors in the list $R_{\text{list}}$ based on current densities at $t = 0$, as per (4.22).
2: Sort $R_{\text{list}}$ in ascending order of TTF samples
3: $\tau_s \leftarrow R_{\text{list}}[1].ttf$
4: for $k = 1 \rightarrow \text{SIZE}(R_{\text{list}})$ do
5: Find $\Delta G_k$, the conductance stamp of $R_{\text{list}}[k]$, as per (3.3).
6: $\{v_k, \text{grid\_singular}\} \leftarrow \text{WB}(v_0, \Delta G_k, L_G, U_G, Z, H, W_{inv}, y, \gamma, k)$
7: if $\text{grid\_singular} = 1$ then
8: $\tau_m \leftarrow R_{\text{list}}[k].ttf$  \hspace{1cm} $\triangleright$ grid failed due to singularity
9: return  \hspace{1cm} $\triangleright$ return to callee
10: end if
11: for $q = 1 \rightarrow \text{SIZE}(v_k)$ do
12: if $v_k[q] > v_{th}[q]$ then
13: $\tau_m \leftarrow R_{\text{list}}[k].ttf$  \hspace{1cm} $\triangleright$ grid failed because $v(t) > v_{th}$
14: return  \hspace{1cm} $\triangleright$ return to callee
15: end if
16: end for
17: if $\texttt{enable\_ttf\_update} = 1$ then
18: Find $\pi(R_{\text{list}}[k])$ as per definitions (5.3).  \hspace{1cm} $\triangleright$ find region of impact
19: Update the TTFs of all the resistors in the set $\pi(R_{\text{list}}[k])$ as outlined in section 4.3.5.
20: Sort $R_{\text{list}}$ from $k + 1 \rightarrow \text{SIZE}(R_{\text{list}})$ in ascending order of TTF samples
21: end if
22: end if

Factors are:

1. **Power grid configuration**, in terms of $\text{max}(v_0)$ and metal temperature $T_m$.
2. **Property of metal used**, in terms of Black’s constant $A$, Activation Energy $E_a$, resistivity of metal $\rho$ and Blech product $\beta_c$.
3. **Choice of failure mechanism**, in terms of current exponent $n$
4. **Effect of statistical parameters**, quantified by standard deviation of lognormal $\sigma_{\ln}$

In all the experiments conducted so far, these factors/ constants were given a specific value, for which the gain ratio was found to be 3-4 for all grids. However, the gain-ratio will be affected if the constants were given different values. In this section, we investigate the effect of varying these constants on the gain-ratio. The test grid used for this study is IBMPG2, but same conclusions can be drawn for other grids as well.

Figures 5.3, 5.4 and 5.5 depict the results of the experiments. Changing Black’s constant $A$, Activation energy $E_a$ and Temperature of metal $T_m$ do not have any effect on the gain ratio.
because if only $A$, $E_a$ and/or $T_m$ are changed and everything else is kept same, the TTFs of all interconnects are scaled linearly, which neither affects the sequence of failure of interconnects nor the ratio $\mu_m/\mu_s$. On the other hand, increase in standard deviation of lognormal, $\sigma_{ln}$, increases the variance of TTFs, so that the difference in TTFs of consecutive interconnects in the failing sequence increases. Thus, the gain ratio increases. It is observed that with increasing current exponent ($n$) the gain ratio decreases and with change in either resistivity $\rho$ of interconnects or Blech product $\beta_c$, the gain ratio fluctuates in a narrow range, but never plummets down drastically. This shows that the mesh model is robust and the increase in EM margin predicted using the mesh model is stable. Finally, as expected, the gain-ratio decreases with increase in $\max(v_0)$, since lesser and lesser interconnect failures are required to exceed $v_{th}$.

### 5.5 Implementation and Experimental Results

The algorithms in section 5.2 and 5.3 have been implemented in C++. The experiments were carried out on both the internal and external grids, using a 2.6 GHz Linux machine with 24 GB of RAM.

For Survival Probability Estimation, $y$ was a user-provided input. As mentioned before, $P_s$ and $P_m$ are estimated with a confidence of 95% ($\alpha = 0.05$) and a maximum relative error of 5% ($\mathcal{E} = 0.05$), which leads to a stopping criteria of $w = 490$ iterations for all the test grids. However, this stopping criterion can be relaxed or made more strict, depending on the need of the user. Note that when only the survival probability of the grid for $y$ years is to be estimated, the actual implementation in C++ differs a little: algorithm `TTF_GRID` (or `TTF_GRID_SELECTIVE`) is modified so that it returns if the last failed interconnect had $\tau > y$ (which means that the grid will definitely survive beyond $y$ for this realization) or if the grid fails before $y$, whichever
Figure 5.4: Variation of gain ratio with property of metal used for power grid IBMPG2

Figure 5.5: Variation of gain ratio with power grid configuration for power grid IBMPG2
Table 5.1: Survival probability Estimation of power grids for user-provided values of \( y \) with 95% confidence and maximum absolute error of ±0.05.

A complete statistical analysis was done for the IBMPG5 grid to estimate the RVs \( T_m \) and \( T_s \). The results are as shown in figure 5.6. A total of 490 grid TTF samples were obtained for each model. Using the goodness-of-fit methods [21], it was found that the samples could be fitted well by a normal distribution (figure 5.6a), so that \( T_m \) and \( T_s \) have a normal distribution. This is further verified by the plots in figure 5.6 b, c and d, which show a very good agreement between the empirical PDF, Reliability and failure rate functions and the actual curves plotted from expressions (2.17), (2.20) and (2.21) with parameter values as obtained from the goodness-of-fit plot. Note that since \( T_m \) and \( T_s \) are shown to have a normal distribution, the use of (2.48) as a stopping criterion to estimate MTF is also verified.
Figure 5.6: Estimated statistics of RVs $T_s$ and $T_m$ for IBMPG5, they are found to have a normal distribution
Table 5.2: Speed and accuracy after using *selective TTF update* approach to estimate power grid MTF, $|\pi(R)|$ is the average size of the set $\pi(R)$ (in terms of % of interconnects) over all failures when $\delta^*_v = 10^{-3}V$ and $\epsilon^+$ is the percentage error in MTF estimation with respect to the original approach.

<table>
<thead>
<tr>
<th>Grid Name</th>
<th>max($v_0$) ($% v_{dd}$)</th>
<th>Mesh Model with <em>selective updates</em>, $\delta^*_v = 10^{-3}V$</th>
<th>Comparision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\mu_m^+$ (yrs)</td>
<td>#MC $w_m$ #Fails $k_f$ Time/Fail $\bar{t}<em>f$ Time/MC $\bar{t}</em>{mc}$ CPU-Time $t_{cpu}$ $</td>
<td>\pi(R)</td>
</tr>
<tr>
<td>IBMPG2</td>
<td>7.0%</td>
<td>34.17 51 19.92 0.18s 3.61s 3.07m 21.77%</td>
<td>↑ 0.32% 1.62x 1.62x</td>
</tr>
<tr>
<td>IBMPG3</td>
<td>6.8%</td>
<td>39.62 78 45.26 2.18s 1.64m 2.14h 6.50%</td>
<td>↓ 0.97% 1.39x 1.59x</td>
</tr>
<tr>
<td>IBMPG4</td>
<td>7.4%</td>
<td>35.70 92 37.86 2.47s 1.56m 2.38h 7.82%</td>
<td>↓ 5.18% 1.55x 0.95x</td>
</tr>
<tr>
<td>IBMPG5</td>
<td>4.8%</td>
<td>31.97 96 17.26 0.84s 14.52s 23.23m 15.47%</td>
<td>↓ 1.16% 1.89x 1.76x</td>
</tr>
<tr>
<td>IBMPG6</td>
<td>5.7%</td>
<td>43.45 58 34.33 1.02s 35.05s 33.88m 12.95%</td>
<td>↑ 3.63% 3.22x 4.22x</td>
</tr>
<tr>
<td>IBMPGNEW1</td>
<td>9.4%</td>
<td>59.12 108 63.91 1.31s 1.39m 2.50h 2.82%</td>
<td>↓ 1.41% 2.11x 1.71x</td>
</tr>
<tr>
<td>IBMPGNEW2</td>
<td>7.7%</td>
<td>39.50 73 42.26 4.02s 2.83m 3.44h 7.09%</td>
<td>↓ 0.36% 1.33x 1.36x</td>
</tr>
<tr>
<td>G1</td>
<td>3.6%</td>
<td>33.88 51 24.10 0.08s 2.03s 1.72m 0.081%</td>
<td>↑ 0.73% 2.50x 4.27x</td>
</tr>
<tr>
<td>G2</td>
<td>4.6%</td>
<td>33.75 93 34.28 0.31s 10.62s 16.47m 0.043%</td>
<td>↑ 0.14% 1.79x 1.47x</td>
</tr>
<tr>
<td>G3</td>
<td>5.0%</td>
<td>36.56 69 45.09 0.43s 19.29s 22.19m 0.025%</td>
<td>↓ 3.46% 2.57x 3.46x</td>
</tr>
<tr>
<td>G4</td>
<td>4.6%</td>
<td>49.34 68 92.13 1.12s 1.72m 1.95h 0.015%</td>
<td>↓ 1.17% 2.69x 2.34x</td>
</tr>
<tr>
<td>G5</td>
<td>4.0%</td>
<td>43.98 55 122.07 2.90s 5.90m 5.40h 0.006%</td>
<td>↑ 2.50% 2.02x 1.73x</td>
</tr>
</tbody>
</table>

**Average Relative error and speed-up**: 1.75% 2.05x 2.19x
Also, since we have a large sample size, we can have a better estimate of \(E[T_s]\) and \(E[T_m]\) for IBMPG5. From the sample mean, \(E[T_m] = 31.6\) yrs and \(E[T_s] = 10\) yrs. Using (2.48), we can arrive at:

\[
\sqrt{w \frac{\bar{X}_w}{s_w}} = \frac{z_{\alpha/2}}{\epsilon/(1 + \epsilon)} \tag{5.4}
\]

The LHS of (5.4) is known, because in this case we already have the samples. Hence, we can choose a specific \(\alpha\) and find the corresponding \(\epsilon\). For both the series and mesh model, we choose \(\alpha = 0.001\) (99.9% confidence), which gives the maximum relative error to be 5.09% and 3.78% respectively. Thus, it can be said, with 99.9% confidence, that

\[
E[T_s] = 10 \pm 0.5\text{ yrs} \quad \text{and} \quad E[T_m] = 31.6 \pm 1.2\text{ yrs} \tag{5.5}
\]

Interestingly, the estimated MTF for series and mesh model, as reported in Table 4.2 is 9.96 yrs and 32.34 yrs, respectively, which falls within the range of (5.5).

Table 5.2 shows the speed and accuracy of the proposed selective TTF-update approach. A user specified value for \(\delta_v^*\) was used for all the test grids \((\delta_v^* = 10^{-3} V)\), and the other parameters are the same as given in section 4.5. The run-time and accuracy is compared with the original approach where for each interconnect failure, the TTFs of all surviving interconnects are updated (as reported in table 4.2). The percentage error is calculated as:

\[
\varepsilon^+ \triangleq \frac{(\mu_m^+ - \mu_m^*) \times 100}{\mu_m^*} \tag{5.6}
\]

The results clearly show that with selective updates, we are able to achieve more than 2x speed-up in both \(T_f\) and \(t_{cpu}\) with small loss in accuracy. The maximum percentage error incurred in MTF estimation is 5.18%, with the average error over all grids being only 1.75%, which is very less when compared to the error in secondary approach (see table 4.3). From figure 5.2, we can observe that for a given value of \(\delta_v^*, |\pi(\mathcal{R})|\) will be larger for external grids as compared to internal grids, which is confirmed from the reported values of \(|\pi(\mathcal{R})|\) in Table 5.2.

For many grids, the speed-up obtained in \(T_f\) for selective updates approach is higher than that of the secondary approach. On first glance this looks inconsistent, the question being: how can updating only node voltage drops be slower than updating node voltage drops and statistics for selective interconnects? The answer lies in scalability of \(t_{f,k,p}\), the time taken to update voltage drops and statistics when the \(k^{th}\) interconnect fails in the \(p^{th}\) monte-carlo iteration. From previous theory, \(t_{f,k,p}\) for selective updates approach scales as:

\[
O(t_{f,k,p}) = O(m^2) + O(m) + O(mk) + O(k^2) + O(1) \tag{5.7}
\]

In the case of secondary approach, the last term is absent. Eq. (5.7) answers the question: in the secondary approach, the length of failure sequences for some grids is so large that it...
over-shadows the speed-up achieved in $\bar{t}_f$ by omitting the $O(1)$ term. Thus, in spite of extra computation involved, the selective update approach is able to achieve greater speed-ups in $\bar{t}_f$. Added to that, it is able to preserve the convergence of the original approach, so that gain in $\bar{t}_f$ is carried over to $t_{cpu}$, resulting in overall speed-up of 2.2x.

Although we obtain a significant speed-up in the selective updates approach, its scalability degrades when compared to the original approach. For the selective updates approach, the total time taken to estimate the MTF, $t_{cpu}$, scales as $O(m^{1.72})$, whereas in the original approach $t_{cpu}$ scales as $O(m^{1.55})$. The scalability of $\bar{t}_f$ and $\bar{t}_{mc}$ also degrade as compared to the original approach, and are found to be $O(m^{1.23})$ and $O(m^{1.65})$, respectively, as shown in figure 5.7. The scalability deteriorates because the convergence becomes slower with increase in grid size, as evident by the increased number of Monte Carlo iterations required to estimate the MTF and longer failure sequences (for large grids) when compared to the original approach.

The constant $\delta_v^*$ controls the speed-up vs accuracy trade-off in MTF estimation using selec-
Figure 5.8: Runtime($t_{cpu}$) and error ($|\varepsilon^+|$) vs. $\delta_v^*$; effect of varying $\delta_v^*$ on $\overline{t_f}$ (average time/failure) and $|\pi(R)|$ (the average cardinality of the set $\pi(R)$). All the results are for IBMPG6.

tive updates. An analysis was performed to study the extent of run-time reduction with $\delta_v^*$ for IBMPG6. The results of the study is shown in figure 5.8. It can be seen from the figures that as $\delta_v^*$ is increased from $10^{-6}$ to $10^{-1}$, $|\pi(R)|$ decreases, which in turn improves $\overline{t_f}$ and $t_{cpu}$, at the cost of increasing the error in estimation. Even though $\overline{t_f}$ decreases with decreasing $|\pi(R)|$, $t_{cpu}$ increases when $\delta v^* \geq 10^{-2}$ due to decrease in convergence (as observed earlier in the secondary approach).
Chapter 6

Conclusions and Future Work

A well-designed power grid in Integrated Circuits (ICs) not only must supply power that is free from fluctuations, but it should also guarantee that its functionality will be reliable for several years. In modern chips, the interconnect structure is arranged in several levels of metalization with thousands of inter-level connections such as via. As the modern designs become more complex and the structural dimensions of electronic interconnects become ever-smaller, the current densities continue to increase. Added to that, interconnect length has been increasing, which further decreases the reliability of interconnects. Due to these scaling trends, the designers are now faced with an ever-reducing margin between the predicted electromigration stress and the stress allowable by electromigration design rules. However, the reduction in margin for the modern designs can also be traced to the pessimism/conservatism built-in the state-of-the-art EM checking tools. A major drawback in the present EM checking tools for power grids is the series system assumption, under which a power grid is assumed to have failed if any of its interconnects fail. Clearly, the series model is highly pessimistic, considering that most modern power grids have a mesh structure.

This necessitates a EM checking tool that moves away from the overly pessimistic series model and takes into account the redundancy of the power grid while estimating its lifetime. In this work, we developed a more realistic grid EM checking and budgeting method that takes the mesh structure of the power grid. We proposed and implemented 1) the mesh model to estimate the MTF and reliability of power grid, 2) a fast and exact approach to update the voltage drops using the Woodbury formula and Banachiewicz-Schur form and 3) a framework to predict the failure statistics of a metal line as it’s effective-EM current density varies over large time-scales. The mesh model accounts for the redundancies in the grid based on the new failure criteria of node voltage drop threshold, and thus obtained a more realistic estimate of grid’s MTF and reliability. The results prove that the current practice of ignoring the redundancy in the grid gives a pessimistic estimate of grid TTF statistics by a factor of 3-4. We also leveraged locality to reduce the number of TTF updates to be performed after an interconnect fails. This spe-
up our MTF estimation engine by $\sim 2x$, enabling us to estimate the MTF of a million node grid under 5.5 hrs. We also verified the stability of the reported increase in the EM margins: even when the input parameters defining the power grid configuration, property of material used etc are changed, the gain ratio never plummets down, we still get at least a 2.5x increase in estimated lifetime by taking the redundancies into account. However, note that since the redundancy model is applicable only to mesh structures and the signal nets hardly have such redundancy (except redundant vias), the 3-4x increase in lifetime predicted for power grids will not translate to a 3-4x increase in the overall chip lifetime.

There are many avenues to further extended this work. A natural extension would be to develop a budgeting framework (akin to SEB for series model) that will enable the chip level reliability to be budgeted among different class of interconnects using the mesh model. Also, since Monte Carlo iterations are highly parallelizable, one can perform these iterations using either the GPUs (CUDA/OpenCL) or cluster computing to greatly improve the run-time. Another proposed future extension could be to use a more realistic TCAD model to simulate degradation in an interconnect due to EM instead of Black’s equation. It will definitely be interesting to solve the reverse problem, which can be stated as follows: given a power grid, the location of current sources and a target MTF, find the (local and global) constraints so that the power grid meets its target lifetime.
Appendices
Appendix A

The Banachiewicz-schur form

Let \( M \in \mathbb{R}^{k \times k} \) be a 2 \( \times \) 2 block matrix:

\[
M = \begin{bmatrix} A & b \\ c^T & d \end{bmatrix}
\] (A.1)

where \( A \in \mathbb{R}^{(k-1) \times (k-1)} \), \( b \in \mathbb{R}^{k-1} \), \( c \in \mathbb{R}^{k-1} \), and \( d \) is a scalar. The Schur-complement of \( A \) in \( M \) is given by [56]:

\[
s = d - c^T A^{-1} b
\] (A.2)

If both \( M \) and \( A \) in (A.1) are non-singular, then \( s \) is non-singular too, i.e. \( s \neq 0 \). This allows writing \( M \) as:

\[
M = \begin{bmatrix} I_{k-1} & 0 \\ c^T A^{-1} & I_1 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & s \end{bmatrix} \begin{bmatrix} I_{k-1} & A^{-1}b \\ 0 & I_1 \end{bmatrix}
\] (A.3)

where \( I_\lambda \) is the identity matrix of order \( \lambda \). The inverse of \( M \) can now be written as [56]:

\[
M^{-1} = \begin{bmatrix} I_{k-1} & -A^{-1}b \\ 0 & I_1 \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & 1/s \end{bmatrix} \begin{bmatrix} I_{k-1} & 0 \\ -c^T A^{-1} & I_1 \end{bmatrix}
\] (A.4)

which can be re-written as:

\[
M^{-1} = \begin{bmatrix} A^{-1} + \frac{A^{-1}bc^T A^{-1}}{s} & -\frac{A^{-1}b^T}{s} \\ -\frac{c^T A^{-1}}{s} & 1/s \end{bmatrix}
\] (A.5)

Equation (A.5) is known as the Banachiewicz-schur form. It expresses \( M^{-1} \) in terms of \( A^{-1} \).
Bibliography


