UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Electromigration-Aware On-Chip Power Grid Design and Optimization

Permalink
https://escholarship.org/uc/item/87t8g16b

Author
Zhou, Han

Publication Date
2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Electromigration-Aware On-Chip Power Grid Design and Optimization

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

Doctor of Philosophy

in

Electrical Engineering

by

Han Zhou

March 2021

Dissertation Committee:

Dr. Sheldon X.-D. Tan, Chairperson
Dr. Hyoseung Kim
Dr. Shaolei Ren

Copyright by
Han Zhou
2021

The Dissertation of Han Zhou is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

This dissertation represents the culmination of work that would not have been possible if not for the support, mentorship, and collaboration from numerous individuals.

Firstly, I would like to express the deepest appreciation to my advisor Dr. Sheldon Tan for his constant guidance and tireless encouragement over the years. His years of experience and his immense wealth of technical and theoretical knowledge have been invaluable to my work and career.

I would also like to thank my committee members, Dr. Hyoseung Kim and Dr. Shaolei Ren for their direction, discussion, and advice for improving my work.

I would like to also thank the members of the VLSI Systems and Computation Lab. Specifically, I would like to thank Taeyoung Kim, Hengyang Zhao, Zeyu Sun, Chase Cook, Shaoyi Peng, Sheriff Sadiqbatcha, Jinwei Zhang, Wentian Jin, Shuyuan Yu, Liang Chen, and Yibo Liu all for their collaboration and advice which has led to this work. Their fellowship has been invaluable and is greatly appreciated.

Lastly, I want to thank my family for their unwavering love and support.

The content of this dissertation is a re-print of the materials that are appeared in the following publications:

- H. Zhou, Y. Sun, Z. Sun, H. Zhao, and S. X.-D. Tan, "Electromigration-Lifetime Constrained Power Grid Optimization Considering Multi-Segment Interconnect Wires", in Proceedings of the 23rd Asia and South Pacific Design Automation Conference (ASP-DAC'18), ICC Jeju, Korea, Jan. 2018, pp. 399-404. (Chapter 3)

- H. Zhou, Z. Sun, S. Sadiqbatcha, N. Chang, and S. X.-D. Tan, "EM-Aware and Lifetime-Constrained Optimization for Multisegment Power Grid Networks", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 4, pp. 940-953, Apr. 2019. (Chapter 3)
- H. Zhou, S. Yu, Z. Sun, and S. X.-D. Tan, "Reliable Power Grid Network Design Framework Considering EM Immortalities for Multi-Segment Wires", in Proceedings of the 25th Asia and South Pacific Design Automation Conference (ASP-DAC'20), Beijing, China, Jan. 2020, pp. 74-79. (Chapter 4)
- H. Zhou, L. Chen, and S. X.-D. Tan, "Robust Power Grid Network Design Considering EM Aging Effects for Multi-Segment Wires", Integration, the VLSI Journal, vol. 77, pp. 38-47, Mar. 2021. (Chapter 4)
- H. Zhou, W. Jin, and S. X.-D. Tan, "Fast Data-Driven EM-Induced IR Drop Prediction and Localized Fixing for On-Chip Power Grid Networks", in Proceedings of the 39th International Conference on Computer-Aided Design (ICCAD'20), Virtual Conference, Nov. 2020, pp. 1-9. (Chapter 5)

To my parents for all the love and support.

ABSTRACT OF THE DISSERTATION

Electromigration-Aware On-Chip Power Grid Design and Optimization

by

Han Zhou

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, March 2021
Dr. Sheldon X.-D. Tan, Chairperson

As technology scales into smaller feature sizes, electromigration (EM) has become a progressively severe reliability challenge due to increased interconnect current densities. It may lead to an increase in wire resistance thus resulting in functional failure of the system. In other words, the reliability of very large scale integrated (VLSI) circuits is increasingly endangered by EM. Therefore, to ensure reliable circuits, it is important to shift today's design flow from the traditional EM verification towards an EM-aware design methodology.

Since power grid wires experience the largest current flows on a chip, they are more susceptible to long-term reliability issues and functional failures. The interconnect wire in power grid networks usually contains multiple segments, which is a multi-segment wire. Generally, the wider the interconnect, the lower the current density and the greater the resistance to EM. On the other hand, during the power grid synthesis in a typical design flow, one important step is to size the wire width of the power grid stripes after the topology of the power supply network has been determined. The area of the power grids is expected to be optimized while EM and excessive IR drop constraints are met.

In this dissertation, several EM-aware power grid wire sizing methods are presented to enhance EM reliability in the traditional optimization approaches and speed up the optimization process. Specifically, first, new power grid network sizing techniques based on the novel EM immortality check methods for general multi-segment interconnects have been proposed. Second, by taking EM aging effects into consideration, we have developed two EM lifetime constrained power grid optimization methods to address the overconservative EM immortality constraint. Furthermore, we have proposed a new data-driven fast EM-induced IR drop estimation framework to accelerate full-chip EM-induced IR drop analysis. Last but not least, we have presented novel approaches for localized and global optimization by utilizing the inherent differentiable power of deep neural networks.

Contents

List of Figures xii

List of Tables xiv

1 Introduction 1
1.1 Motivation . . . 1
1.1.1 EM-aware power grid wire sizing . . . 2
1.1.2 Machine learning accelerated power grid analysis . . . 5
1.2 Contributions . . . 6
1.3 Organization . . . 10

2 Preliminaries of EM physics and power grids analysis 11
2.1 Electromigration fundamentals and existing models . . . 11
2.1.1 Electromigration fundamentals . . . 11
2.1.2 Physics-based three-phase EM models . . . 14
2.2 Power grid network models . . . 18
2.2.1 Styles of on-chip power distribution networks . . . 18
2.2.2 EM in power grids . . . 19
2.2.3 Modeling assumptions . . . 19
2.2.4 Power grid analysis . . . 20
2.3 Full-chip EM-induced IR drop analysis . . . 21

3 EM immortality and lifetime constrained power grid wire sizing 25
3.1 EM immortality check and lifetime estimation for multi-segment wires . . . 25
3.1.1 Steady-state EM-induced stress modeling . . . 25
3.1.2 Transient EM-induced stress estimation . . . 27
3.2 EM immortality constrained optimization for multi-segment interconnects . 31
3.2.1 Problem formulation . . . 31
3.2.2 Relaxed two-step sequence of linear programming solution . . . 34
3.2.3 EM immortality constrained power grid optimization . . . 36
3.3 EM immortality constrained power grid wire sizing with reservoir insertion 38
3.4 EM lifetime constrained power grid optimization . . . 43

3.4.1 EM lifetime constrained power grid optimization flow 44
3.4.2 Transient EM analysis for a multi-segment interconnect 47
3.5 Experimental results and discussions 50
3.5.1 Experiment setup 50
3.5.2 EM immortality constrained power grid optimization results 51
3.5.3 EM immortality constrained power grid optimization with reservoir insertion results. 54
3.5.4 EM lifetime constrained power grid optimization results 56
3.6 Summary 58

4 Robust power grid network design considering EM aging effects 59
4.1 Saturation volume-based EM immortality check 59
4.2 New EM immortality constrained optimization for multi-segment interconnects 62
4.2.1 New EM immortality criteria 63
4.2.2 Problem formulation 65
4.2.3 Relaxed two-step sequence of linear programming solution 67
4.3 Comprehensive power grid optimization strategies 69
4.3.1 New EM immortality constrained power grid optimization 69
4.3.2 EM immortality constrained power grid optimization with wire pre-sizing 72
4.4 IR drop constrained power grid optimization 73
4.4.1 EM lifetime constrained power grid optimization 73
4.4.2 Sensitivity-based localized power grids fixing 74
4.5 Experimental results and discussions 79
4.5.1 New EM immortality constrained power grid optimization results . . 79
4.5.2 EM immortality constrained power grid optimization results with wire pre-sizing 81
4.5.3 EM lifetime constrained power grid optimization results 82
4.5.4 Sensitivity-based localized power grid fixing results 82
4.6 Summary 84

5 GridNet: Fast data-driven EM-induced IR drop prediction and localized fixing for on-chip power grid networks 86
5.1 Related works 86
5.1.1 Machine learning-based IR drop analysis and estimation 86
5.1.2 EM-induced IR drop analysis and fixing works 87
5.2 Fast data-driven incremental EM-induced IR drop prediction 89
5.2.1 Overall workflow of the *GridNet* framework 89
5.2.2 Feature selections for *GridNet* 91
5.2.3 Training data preprocessing and representation 92
5.2.4 *GridNet* architecture 95
5.2.5 Fast sensitivity calculation using the automatic differentiation in DNNs 97
5.3 Fast layout fixing for EM-induced IR failures 98
5.3.1 Fast localized IR drop fixing 99
5.3.2 Sensitivity-based localized IR drop fixing 100

5.4 Experimental results and discussions 101
5.4.1 Experiment setup 101
5.4.2 EM-induced IR drop prediction results 102
5.4.3 Fast IR drop fixing results 110
5.4.4 Sensitivity-based localized fixing results 111
5.5 Summary 112

6 GridNetOpt: Fast full-chip EM-aware IR drop constrained power grid optimization via deep neural networks 114
6.1 Related works 114
6.2 New EM-induced voltage constrained optimization problem 115
6.2.1 Problem formulation 115
6.2.2 Penalty method 117
6.3 Optimization strategies 119
6.3.1 Conjugate gradient method 119
6.3.2 DNN-based fast EM-induced IR drop estimation 121
6.3.3 Gradient calculation for the objective function 122
6.3.4 Gradient calculation via adjoint network method 122
6.3.5 Fast gradient calculation via deep neural networks 124
6.4 Experimental results and discussions 126
6.5 Summary 129

7 Conclusions 130
7.1 Summary of contributions 130
7.1.1 EM immortality and lifetime constrained power grid wire sizing . . . 130
7.1.2 Robust power grid network design considering EM aging effects . . . 131
7.1.3 Fast data-driven EM-induced IR drop prediction and localized fixing for on-chip power grid network 132
7.1.4 Fast full-chip EM-aware IR drop constrained power grid optimization via deep neural networks 133

Bibliography 135

List of Figures

2.1 Example of a multi-segment wire. 13
2.2 (a) The three-phase EM model; (b) the resistance change over time. 14
2.3 Side-view of void formation: (a) void in a via-above line (early failure mode); (b) void in a via-below line (later failure mode). 17
2.4 A small portion of a typical power grid network [43]. 19
2.5 Equivalent circuit of the small portion of a typical power grid. 21

3.1 Interconnect example of a straight 3-terminal wire for EM analysis. 27
3.2 Example of a straight multi-segment wire. 28
3.3 The 3D view of a two-segment wire with a reservoir. 39
3.4 The impact of length and width of the reservoir on the TTF of a wire. . . . 40
3.5 Inserted reservoirs in the power grid network. 41
3.6 DRC check for reservoir segment placement. 43
3.7 Flowchart of the EM lifetime constrained power grid optimization process. . 45
3.8 The electrical impact of different failure mechanisms on the interconnect wires: (a) early failure mode; (b) late failure mode. 50

4.1 (a) A two-segment wire; (b) stress integration area of a two-segment wire. . 61
4.2 Void formation examples in a power grid network. 63
4.3 Saturation volume-based EM immortality check for power grid networks. . . 71

5.1 *GridNet* training flow. 90
5.2 *GridNet* prediction flow and grids fix flow with *GridNet*. 91
5.3 (a) Power and ground networks of Cortex-M0 DesignStart; (b) voltage drop map of the power network of (a). 93
5.4 Compact IR drop image of power grid networks of (a) Design 2: 4k nodes; (b) Design 3: 16k nodes. 94
5.5 The CGAN architecture for EM-induced voltage prediction. 96
5.6 Comparison of inference results from *GridNet* and *EMspice*: (a) IR drop distribution at different years; (b) increased IR drop due to EM; (c) predicted voltage error. 105
5.7 Predicted IR drops versus the baseline of (a) Design 1.1; (b) Design 1.2. . . 107

5.8 Error histogram of (a) Design 1.1; (b) Design 1.2. 107
5.9 Predicted lifetime versus baseline. 108
5.10 Prediction and simulation speed comparison between *GridNet* and *EMspice*. 109

List of Tables

3.1 Parameters used in the optimization process. 51
3.2 EM immortality constrained power grid optimization results for IBM power grid networks. 52
3.3 EM immortality constrained power grid optimization with reservoir insertion results for IBM power grid networks. 55
3.4 EM lifetime constrained power grid networks for self-generated power grid networks. 56

4.1 EM immortality and lifetime constrained power grid optimization considering void saturation volume results. 79
4.2 Sensitivity-based localized fixing results. 83

5.1 Prediction results of different designs. 103
5.2 Error statistics of Design 1.1 and Design 1.2. 104
5.3 Fast IR drop fixing results. 110
5.4 Sensitivity-based localized fixing results. 111

6.1 Comparison of *GridNetOpt* against adjoint network method. 128

Chapter 1

Introduction

1.1 Motivation

On-chip power supply or power-ground network provides power to the circuit modules in a chip from external power supplies. It is a crucial backbone for feeding voltage sources to all transistors in cells from top metals on a chip as it affects performance (IR drop, timing) as well as area and cost (routability, layout density, metal stack). As the complexity of power-ground networks increases, methods for efficient analysis and aggressive optimization of these networks become essential [44].

As technology scales into smaller features with increasing current densities, electromigration (EM) remains one of the top killers for copper-based interconnects in current and near-future advanced very large scale integrated (VLSI) circuit technologies. An effective scaling model predicted that the EM lifetime decreases by half for each new generation [1]. However, the latest international roadmap for devices and systems (IRDS) [3] indicates that interconnect resistance has now entered an exponential increase regime because of non-ideal

scaling. As a result, the EM-related aging and reliability will become worse for current 5nm and below technologies.

1.1.1 EM-aware power grid wire sizing

For practical VLSI chips, since power grid wires experience the largest current flows on a chip, they are more susceptible to long-term reliability issues and functional failures. These reliability issues and failures typically come from electromigration, excessive IR drops, and ΔI (Ldi/dt) noise along with the back end of line time-dependent dielectric breakdown (TDDB) [9, 7, 75, 14].

From EM perspective, interconnect wire in a power grid network is usually a tree structure containing multiple segments. A multi-segment interconnect tree consists of continuously connected high-conductivity metal within one layer of metallization. Due to EM aging and failure effects, the voltage drop of power grids that meets the design requirement at the design time may get worse and lead to timing violations as time goes by. This introduces additional challenges for designing robust power grid networks to satisfy the demanding design requirements.

An important step for power supply synthesis in the typical EDA design flow is to size the wire width of the power grid stripes after the topology of the power supply network has been determined, so that the minimum amount of chip area will be used while avoiding potential reliability failures due to electromigration and excessive IR drops. The area of the power grids can be optimized while EM and excessive IR drop constraints are met. Many practical solutions have been investigated for the power grid network optimization actively over the past years, based on nonlinear or sequence of linear programming (SLP)

methods [18, 17, 16, 24, 56, 57, 61]. To satisfy the EM reliability, most research efforts adopted the current density of individual wires as the constraint, which is mainly based on the highly conservative Black's EM model [9]. Empirical observation has shown that a wire's mean time to failure depends on the current density through its cross-section and limiting this current density can control EM [44].

However, recent studies and experimental data show that the time to failure predicted by Black's EM model is too conservative and thus more accurate physics-based EM models have been proposed [45, 33]. More importantly, practical VLSI interconnects, especially the global networks such as power supply and clock networks, contain a lot of multi-segment wires. Studies show that the current-induced stress developed in those segments is not independent [59, 29, 55]. In other words, the hydrostatic stress in multi-segment interconnects is coupled and the EM failure condition of the entire interconnect must be considered as a whole [47, 33, 41, 12, 13, 55, 19, 54]. Since the parameter change of one wire segment may affect the EM stress condition of another wire segment in the same interconnect tree, it provides more powerful potential to optimize a power grid network subject to EM constraints for multi-segment wires compared with the traditional current density-based methods. By comparing the so-called *EM voltage* against the *critical EM voltage*, a fast EM immortality check method [50] checks all the wire segments to determine immortality. However, this new multi-segment EM immortality check has not been further applied to optimize power grid networks, which will be one of the focuses of this work.

With further development of EM research, the above-mentioned voltage-based EM immortality check has shown limitations: void nucleation is not allowed. For a multi-

segment interconnect, once a void is formed, it will start to grow. However, the void growth will stop when the compressive stress in the remaining wires reaches the steady-state (the back force from the compressive stress is equal to the force for void growth). The void volume at this time is called *saturation volume*. If the saturation volume of a void is smaller than the so-called *critical volume*, then the wire will still be considered immortal. The *critical volume* is defined as the volume (or *critical area* in the two-dimensional case) that the void must reach or exceed to completely block the current flow in the copper interconnect and shunt it to the metal liner [51, 52]. Therefore, the conservative nature of the power grid optimization process can be further relaxed if this saturation volume effect is considered, which will be another focus of this work.

Another important issue with the existing power grid network optimization methods is that they fail to consider the aging effects. In other words, they can hardly optimize a power grid network that has nucleated wires or even completely failed wires at the target lifetime, even though the whole power grid network may still work. The difficulty for such optimization lies in the need for accurate time to failure estimation and transient wire resistance change estimation methods. With recent advancements in physics-based EM models and numerical analysis techniques such as three-phase EM models [63, 49, 19, 55], it has been possible to perform more accurate time to failure estimation for multi-segment interconnects.

Furthermore, all the existing power grid optimization methods mainly consider wire sizing which impacts the lifetime of wires, rather than take other more effective optimizing knobs into account, i.e., adding or relocating reservoirs or sinks which belong to the

interconnect. These approaches benefit from complicated physics-based EM models, which have become possible with the recent advanced techniques.

1.1.2 Machine learning accelerated power grid analysis

As technology advances, interconnect resistance of proportionally scaled wires has increased significantly, making IR drops a major factor in the performance of power grid networks. Since the stress and IR drops are interrelated, if the EM constraint or EM aging effect can be directly integrated into the IR drop constraint, then the optimization formulation for power grids would become much simpler.

In the past few years, research efforts on physics-based EM analysis for power grid networks have been explored, especially a coupled time-varying EM and IR drop analysis method [53]. However, all the methods need to solve the partial differential equations (PDE) of hydrostatic stress for multi-segment interconnect wires, which is very time-consuming for full-chip analysis.

On-chip power grid networks typically go through many iterations between power grid designs, IR analysis, and floorplan/placement updates before the final EM sign-off. Typically power grid networks at late design stage are well designed and sized to meet the IR drop requirement, both static and dynamic. But IR drops at a few nodes may still fail at the target lifetime due to EM aging and failure process. At this stage, engineering change order revisions or updates for the power grid layout will be carried out.

Hence, there is a need for fast EM-induced IR drop estimation for power grid networks. Such analysis will be different from the traditional fast IR drop analysis meth-

ods [39, 25, 30, 66] as those methods are mainly aimed at accelerating the design time dynamic IR drop analysis.

On the other hand, deep neural networks (DNN) have propelled an evolution in machine learning fields and redefined many existing applications with new human-level AI capabilities. They have been applied to many cognitive applications such as visual object recognition, object detection, speech recognition, natural language understanding, etc., due to dramatic accuracy improvements in those tasks [37]. Recently, generative adversarial network (GAN) [28] gained increasing attention as it can learn features (latent representation) without extensively annotated training data.

GAN-based methods have been applied to VLSI physical designs such as various noise map generation to facilitate the sensor placement [40], layout lithography analysis [68], sub-resolution assist feature generation [5], analog layout well generation [67]. However, these GAN-based designs and analyses are mainly targeted for the statistical and static image generations (analysis). Few works have been explored to learn time-series data.

1.2 Contributions

The work presented in this manuscript presents several contributions in the area of EM-aware on-chip power grid network design and optimization:

- For EM immortality constrained power grid optimization, a new power grid network sizing technique based on the voltage-based EM immortality check method for general multi-segment interconnect wires is proposed. This power grid network optimization problem subject to the voltage IR drop and new EM constraint can still be formu-

lated as an efficient SLP problem, where the optimization is carried out in two linear programming phases in each iteration. The new optimization ensures that none of the EM wires fail as long as all the constraints are satisfied, which is not possible with the existing current density constrained optimization methods. Then, by taking advantage of a new saturation volume estimation method for general multi-segment interconnects, we illustrate that some power grids with nucleated wires that cannot be optimized using the voltage-based EM immortality constraints can be further optimized. In other words, we consider two EM immortality constraints: *EM nucleation phase immortality* and *EM incubation phase immortality*. Then we show that both EM immortality conditions can be naturally integrated into the linear programming-based power grid optimization framework, which remains the most efficient power grid optimization method. In addition, to mitigate the overly conservative nature of the optimization formulation, we propose to insert proper reservoirs or size up failed wires to meet one of the immorality conditions subject to the design rules so that the programming-based optimization can be carried out.

- For EM lifetime constrained power grid optimization, two power grid network sizing techniques from different perspectives have been proposed. To address the overconservative issue of EM immortality constraint, EM aging effects are taken into consideration: one is for interconnect wires and the other is for the entire power grid network. The first method aims to enforce the target lifetime for all the wires. It is by means of considering the aging effects of interconnects at the target lifetime, which is based on the physics-based EM assessment technique. Specifically, this method

allows a part of short-lived wires to fail and optimizes the rest of the wires in the presence of wire resistance increase. In this way, the original power grid networks can be optimized and the resulting networks become more robust and aging-aware over the lifetime of the chip. In the second method, we consider the EM-induced aging effects over power grid networks for a target lifetime. A large change sensitivity-based optimization scheme is proposed to perform localized power grids fixing. The method, based on the new coupled EM-IR analysis method for power grid networks, may fix EM-induced IR drops in a few minutes.

- For full-chip EM-induced IR drop prediction, we proposed a data-driven fast EM-induced IR drop analysis framework, called *GridNet*, based on the conditional generative adversarial networks. *GridNet* treats the EM-induced time-varying voltage of power grid networks as the time series data images which are conditional outputs of the given electrical features. It can lead to five orders of magnitude speedup over the full-chip coupled EM-IR analysis tool *EMspice*. We show that GAN model can be adopted to learn temporal dynamics in the power grid aging process by using the continuous time as one of the conditions. More importantly, *GridNet* can not only perform fast if-then analysis for EM-induced IR drop estimation but also provide important sensitivity information of node voltage with respect to the wire resistance with marginal cost. It is obtained by leveraging the differentiable feature of DNNs as we can easily extend differentiable loss function of DNNs for the input data as well. For our DNN model, we select resistance of all the wire segments and current sources attached at certain nodes as the electrical features, assuming that the mesh-structured

power grid networks have given voltage sources. Since we capture the IR drop in view of incremental circuit analysis, our model can be used on power grid designs with different workloads, different wire widths and lengths without retraining as long as the main power grid structure and voltage sources remain unchanged.

- By utilizing the inherent differentiable function of DNNs for fast sensitivity computation, a new fast EM-induced IR drop constrained optimization framework, called *GridNetOpt*, is developed for full-chip power grid networks. The new optimization framework first builds an unsupervised DNN-based model for EM-induced IR drop of power grid designs with various wire widths and current workloads. With the DNN-based model (*GridNet*) trained in the first step, the sensitivity information of node voltage with respect to the wire resistance/conductance can be obtained with marginal cost for any power grid designs of the same topology via the auto-differentiation function of the DNN model. First, for localized IR drop fixing, performing a localized fixing on vulnerable power grids with data from *GridNet* at late design stages only takes a few seconds, which is extremely efficient. Second, for global power grid optimization, we frame the area sizing problem of the time-varying power grid network and solve it via the conjugate gradient method, in which the sensitivity information is computed by *GridNet*. Compared with conjugate gradient optimization method using the traditional adjoint network approach, *GridNetOpt* can lead to an order of magnitude speedup.

1.3 Organization

The rest of this dissertation is organized as follows. Chapter 2 provides the background knowledge of physics-based EM models. It then focuses on on-chip power grids modeling and analysis. A full-chip EM-induced IR drop analysis method is also introduced. Chapter 3 presents novel formulation and solution methods of EM immortality constrained and EM lifetime constrained power grid wire sizing problems. A more relaxed EM immortality constrained power grid optimization based on saturation volume is developed in Chapter 4, an IR drop constrained power grid network fixing method is also presented. Chapter 5 investigates a novel optimization scheme employing machine learning techniques, where sensitivity information can be obtained quickly through automatic differentiation operations. Chapter 6 further presents a global power grid optimization method, utilizing the cheap sensitivity information from deep neural networks. Finally, Chapter 7 concludes the dissertation.

Chapter 2

Preliminaries of EM physics and power grids analysis

2.1 Electromigration fundamentals and existing models

2.1.1 Electromigration fundamentals

Electromigration is a major reliability concern for interconnect wires with increasing current densities under advanced technology nodes. It is a physical phenomenon of material migration caused by an electrical field. Wind force, which is produced by current flowing through a conductor, acts in the direction of the current flow and is the primary cause of EM [38]. Atoms (either lattice atoms or impurities) migrate from the cathode to the anode of the metal wire along the trajectory of conducting electrons. Because of the diffusion barriers, atoms cannot escape the metal volume easily. This oriented atomic flow leads to metal depletion at the cathode and accumulation at the anode. During the

migration process, hydrostatic stress is generated inside the embedded metal wire due to momentum transfer between lattice atoms. Void and hillock formations are caused by conducting electrons at the opposite ends of the wire.

The currently employed method for predicting the median time to failure (MTF) is based on an empirical model, Black's equation [9], for simple linear interconnects.

$$MTF = \frac{A}{j^n} \cdot \exp\left(\frac{E_a}{k_B T}\right) \tag{2.1}$$

where MTF is the median time to failure in hours, A is a cross-sectional-dependent constant, j is the current density, E_a is the EM activation energy, k_B is the Boltzmann's constant, and T is the temperature. The variable n is to allow the model to be applied to different types of dominant failure mechanisms.

The main disadvantage is that this equation is designed for linear interconnects and cannot successfully be applied to entire wires with changes in direction or changes in layer. Besides, calculating the MTF of individual branches characterized by known current densities and temperatures is the subject of growing criticism as it does not scale very well over a wide range of current densities [38, 55].

On the other hand, the Blech limit or Blech product [10] is used for the filtering out of immortal branches.

$$(jL) \leq (jL)_{crit} = \frac{\Omega \sigma_{crit}}{eZ\rho} \tag{2.2}$$

where L is the wire branch length, Ω is the atomic volume, e is the electron charge, eZ is the effective charge of the migrating atoms, ρ is the wire electrical resistivity, and σ_{crit} is the critical stress needed for void or hillock formation as a precursor to failure.

It should be mentioned that the Blech limit is valid only for a single wire segment or branch within the diffusion blocking boundaries. It does not work for multi-segment interconnect trees. These trees are terminated by diffusion barriers at vias and contacts and can have more than one terminating segment, as shown in Fig. 2.1. Existing EM modeling and analysis techniques mainly focus on the simple straight line interconnect with two line-end terminals. However, clock and power grid networks in a practical integrated circuit layout contain many such interconnect trees. The EM effects in these segments are not independent and have to be considered simultaneously.

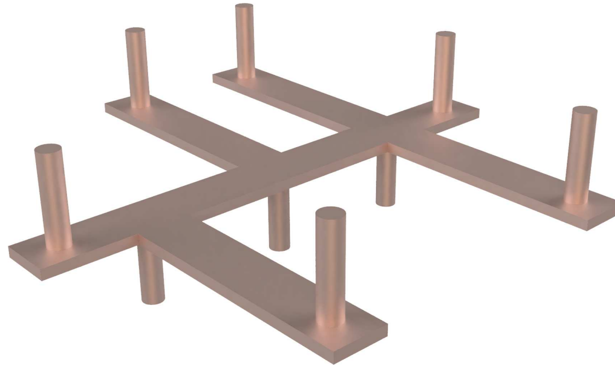


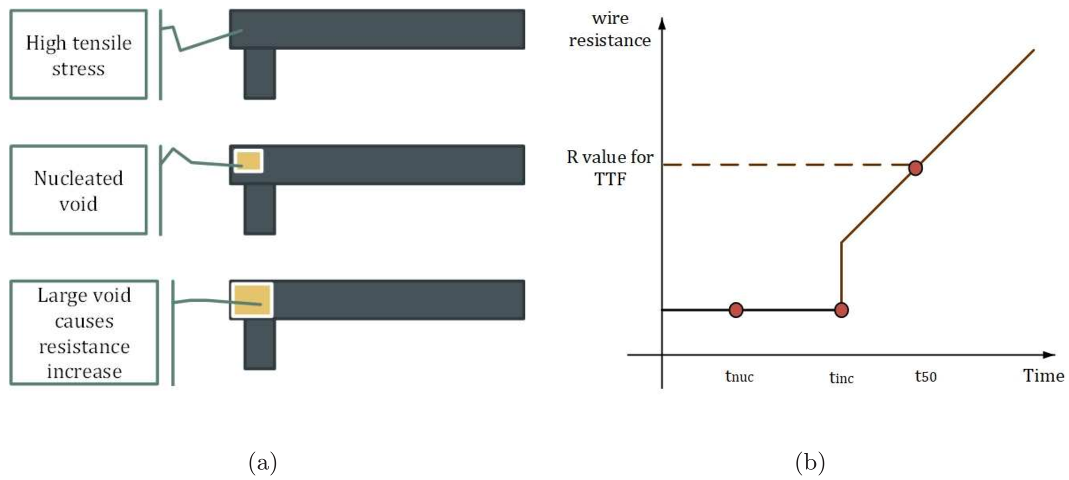
Figure 2.1: Example of a multi-segment wire.

To mitigate the major flaws in the Black-Blech models, several physics-based EM models have been developed recently [34, 46]. Void is responsible for time-dependent resistance degradation of interconnect wire, in these models, void nucleation and void evolution are explicitly characterized.

2.1.2 Physics-based three-phase EM models

As shown in Fig. 2.2, the EM failure process consists of three important phases: void nucleation phase, void incubation phase, and void growth phase. In the nucleation phase, the stress at the cathode end of the interconnect wire increases. When it reaches critical stress, a void will be nucleated. The time to reach the critical stress is called nucleation time (t_{nuc}). After the nucleation phase, the void starts to grow (t_{inc}) and eventually leads to wire failure after a period of time (t_{growth}). The time to failure (TTF) or lifetime of the wire can be described as

$$TTF = t_{life} = t_{nuc} + t_{inc} + t_{growth} \tag{2.3}$$



(a) (b)

Figure 2.2: (a) The three-phase EM model; (b) the resistance change over time.

Nucleation phase

When current is applied to a wire, the tensile stress will start to increase over time at the cathode end of the wire ($x = 0$). When the tensile stress reaches the critical stress σ_{crit}, the void will be nucleated.

As metal wires in integrated circuits are confined, i.e., encapsulated, by dielectric material, every material movement annihilates and/or generates vacancies. There are different mechanical stress generation and evolution models to explain this process. Korhonen's model [36] is one of them and it describes the stress through a one-dimensional equation as follows

$$\frac{\partial \sigma(x,t)}{\partial t} = \frac{\partial}{\partial x}[\kappa(\frac{\partial \sigma(x,t)}{\partial x} + \Gamma)] \tag{2.4}$$

where σ is the hydrostatic stress, x is coordinate along the line, t is time, $\kappa = \frac{D_a B \Omega}{k_B T}$, and $\Gamma = \frac{eZ}{\Omega}\rho j$. Here, Da is the atomic diffusivity, B is the applicable modulus, Ω is atomic volume, k_B is Boltzmann constant, T is absolute temperature, Z is the effective charge number, e is the electron charge and j is the current density.

Incubation phase

In this phase, the void is nucleated, but its size is not significant enough and the void cannot fully cover the cross-section of the via or the wire. Hence the resistance of the interconnect remains almost the same. The incubation time t_{inc} can be expressed as

$$t_{inc} = \frac{\Delta L_{crit}}{v_d} \tag{2.5}$$

where ΔL_{crit} is the length of critical void size and v_d is the void growth rate. v_d is given by the mobility (Da/k_BT) times the electromigration driving force F_e [31], then we have

$$v_d = \frac{D_a}{k_BT}F_e = \frac{D_a}{k_BT}eZ\rho j \tag{2.6}$$

It is noted that in this phase, the approximated drift velocity v_d is proportional to the current density j and thus t_{inc} is related to j^{-1}. The incubation time also reflects the technology scaling impacts on EM. As the feature size of each generation of technology shrinks, which is closely related to minimum via or wire sizes, the EM model is scalable and predictable for technology scaling in addition to scalability in stressing current density.

Growth phase

A copper (Cu) wire structure is confined by the highly resistive barrier layer (such as Ta/TaN) with a capping layer (such as SiN) at the top. The capping layer is fabricated with dielectrics with very high resistance, so it does not shunt current flow.

At the beginning of the growth phase, the void reaches the critical void size and blocks the cross-section above the via. Then the current starts to flow over the liner or barrier layers. Since the liner resistivity is much larger than copper and the liner is very thin, current density of the liner can be very high. In consequence, the high current density and high resistance lead to significant joule heating. As shown in Fig. 2.2(b), joule heating causes the temperature to rise, which brings about a resistance jump at the beginning.

After this jump, the resistance will increase linearly. For a confined Cu wire with a highly resistive barrier layer such as Ta surrounding the three sides of the wire, using the

same drift velocity, the wire resistance change can be approximately described as

$$\Delta r\left(t\right) = v_d\left(t_{growth}\right)\left[\frac{\rho_{Ta}}{h_{Ta}\left(2H+W\right)} - \frac{\rho_{Cu}}{HW}\right] \tag{2.7}$$

where ρ_{Ta} and ρ_{Cu} are the resistivities of the barrier material and copper, W is the line width, H is the copper thickness, and h_{Ta} is the barrier layer thickness.

Note that the void can cause either early failure or late failure of the wire [4]. Early failure typically happens in a via-to-via structure as shown in Fig. 2.3(a). When the void forms in a via-above line and reaches critical size [31, 69], the via will be blocked by the void, and thus the connection to the upper layer will also be blocked. Late failure typically happens in a so-called via-below structure as shown in Fig. 2.3(b), when the void forms in a via-below line and reaches critical size, current can still go through the barrier layer and the resistance will increase over time. Statistically, early failure can happen in the via-below structure and late failure can happen in the via-above structure. Although the void can grow at these positions, the possibility is very low.

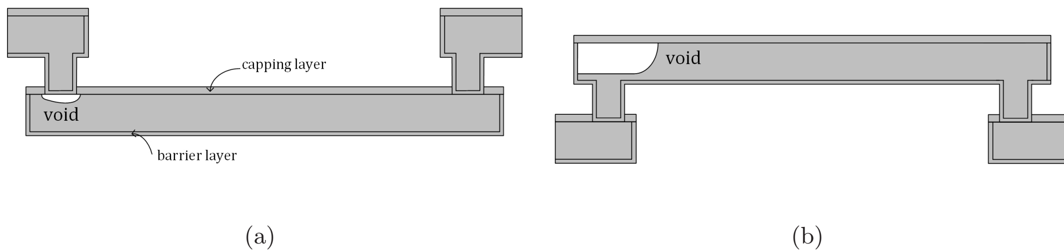


Figure 2.3: Side-view of void formation: (a) void in a via-above line (early failure mode); (b) void in a via-below line (later failure mode).

2.2 Power grid network models

With the VLSI technology scaling, power supply voltage continues to decrease as device number on a chip continues to increase, power network design becomes a challenging task for chips with millions of transistors. The common task in VLSI power network design is to provide enough power lines across the chip to reduce the voltage drops from the power pads to the center of the chip. The voltage drops are mainly caused by the resistance or inductance of the power network metal lines [75]. Generally, the power delivery analysis can be split into two problems, a detailed steady-state problem and a simplified time-domain problem.

2.2.1 Styles of on-chip power distribution networks

Several topological structures are typically used in the design of on-chip power distribution networks. The power network structures range from completely irregular structures to highly regular and uniform structures. The basic topologies include routed networks, mesh networks, grid structured networks, power and ground planes, cascaded power/ground rings, and hybrid-structured networks [35].

Fig. 2.4 shows a typical mesh-structured power/ground network with multi-layer power grids. This example is somewhat idealized. In a realistic design, the mesh is not complete, i.e. some wires may be missing or truncated. Also, the periodicity and density of the wires may vary, so areas of the chip which require less power may have fewer and narrower power grid wires than areas that are more power-hungry.

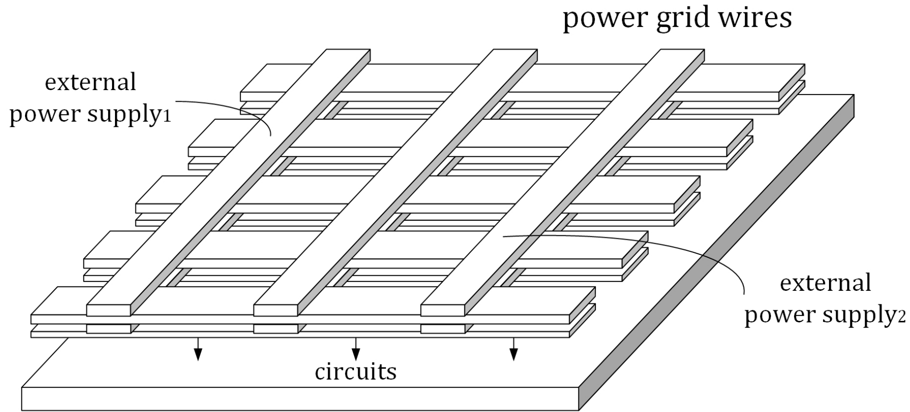


Figure 2.4: A small portion of a typical power grid network [43].

2.2.2 EM in power grids

Electromigration is a serious problem in submicron designs, in most cases, it only affects the power distribution network as the wires carry unidirectional currents. The problem occurs when the power wires are too narrow or an insufficient number of contacts or vias are placed to carry high current density. Current density can be reduced by increasing the size or number of the metal wires. Generally, the use of more metal wires and vias in the power distribution network can help reduce EM failures as well as excessive IR drops. Therefore, enough power wires should be provided in the early design planning stage.

2.2.3 Modeling assumptions

In this section, we will describe the modeling assumptions in this work. The assumptions are used in extracting the equivalent electrical characteristics of the power grids. Fig. 2.5 shows the equivalent circuit of the power grid network in Fig. 2.4.

1. We have a power distribution network containing a power network and a ground network, however, our optimization attention is restricted to power networks, which consists of V_{DD} nets only (also referred to as power grids in the following).
2. We focus on the steady-state (DC) problem, which means we are only interested in the resistance of the power grid networks. Transient behaviors addressed by decoupling capacitor allocation and other optimization techniques are out of the scope of this work.
3. The power grid network is composed of an orthogonal mesh of wires and contains multiple segments/branches, which are multi-segment interconnect wires.
4. To simplify the problem, the circuits are modeled with shorted vias, which means the via resistance is ignored and vias will not be sized. The first reason is that vias cannot be sized under most of the design rules. Second, the void typically forms in a wire segment rather than the via unless it is due to initial thermomechanical stress from the fabrication process. Furthermore, the via resistance is quite small compared to the wire resistance, thus it can be ignored.

2.2.4 Power grid analysis

Analyzing the power grids accurately requires solving a system of linear equations with many variables. The power grid networks in this work are modeled as linear resistance networks with known voltage and current sources.

For a power grid network with n nodes, to obtain the initial nodal voltages and branch currents, the following linear equation which is formulated through modified nodal

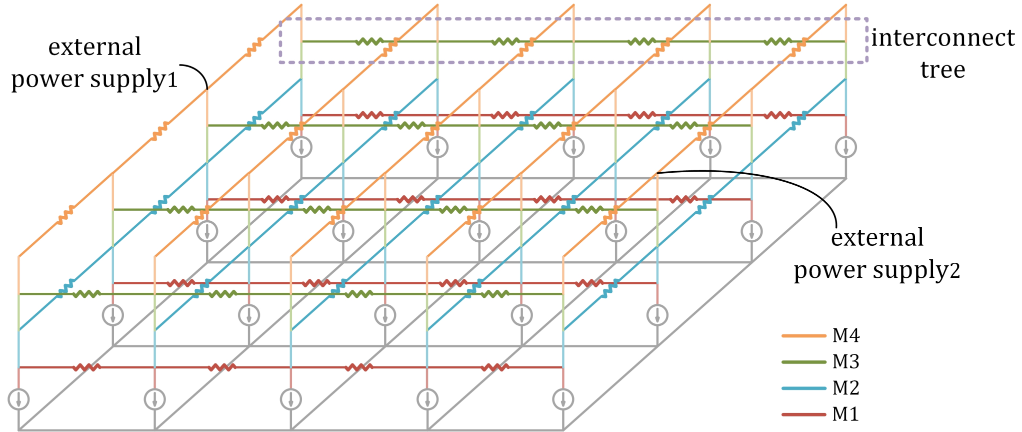


Figure 2.5: Equivalent circuit of the small portion of a typical power grid.

analysis (MNA) approach must be solved first:

$$G(t) \times v(t) = I(t) \tag{2.8}$$

where $G(t)$ is an $n \times n$ conductance matrix, $I(t)$ is the vector of independent current sources, and $v(t)$ is the corresponding vector of voltages at the nodes. In our problem, each EM mortal wire will start to change its resistance value upon entering into the growth phase, therefore, the power grid system is a time-varying system. The time scale is the EM time scale, which can be months or years.

2.3 Full-chip EM-induced IR drop analysis

EM aging process typically leads to resistance increase or even open-wire segments over time. However, for on-chip mesh-structured power grid networks, due to its inherent redundancy, a few wire failures may not immediately result in a significant IR drop increase. But as more wires nucleate, the IR drop will eventually lead to timing violations. As a result, the power grid networks become time-varying networks with time-varying IR drops due to

the EM-induced aging process [33, 13, 49, 48]. On the other hand, the failed wire segments alter the current distributions of all the interconnect wires, which may further accelerate the failure process. Hence, one has to consider the interplay between the two physics: electrical characteristics and hydrostatic stress in the interconnect wires.

EMspice [53, 2] is a coupled electrical and stress analysis method on full-chip power grids, which represents the latest development for EM-induced IR drop analysis. This method will be used as the baseline for the EM-induced IR drop prediction method presented in Chapter 5. *EMspice* takes synthesized power network from Synopsys IC Compiler and then tells which wires will fail, how their resistance will change and how the resulting IR drop of the power grids will increase over the aging time.

For a specific design, the entire EM check in *EMspice* consists of several steps. In the first step, the power grid information is constructed from Synopsys IC Compiler during the physical synthesis process. Next, the power grid and the corresponding branch currents are passed to the EM immortality filter to remove all the immortal wires. This step considers the wire immortality for both nucleation and incubation phases. Thirdly, it solves the stress and IR drop of interconnect wires in a coupled way. The coupled solver consists of a finite difference time domain (FDTD) solver for EM stress [13, 19] and a linear network DC solver for IR drop. Lastly, all information will be fed into the graphical user interface (GUI) for interactive user analysis.

In the third step, the coupled FDTD EM solver and linear network IR drop analysis can be described as

$$
\begin{aligned}
\mathbf{C}\dot{\sigma}(t) &= \mathbf{A}\sigma(t) + \mathbf{P}I(t),\\
\mathcal{V}_v(t) &= \int_{\Omega_L} \frac{\sigma(t)}{B} d\mathcal{V},\\
\mathbf{M}(t) \times u(t) &= \mathbf{P}I(t),\\
\sigma(0) &= [\sigma_1(0), \sigma_2(0), ..., \sigma_n(0)] \ , at \ t = 0
\end{aligned}
\tag{2.9}
$$

where $\mathbf{M}(t)$ is a time-varying admittance matrix of the power grid network due to the fact that wire resistance changes with EM aging effects, $\mathbf{P}$ is a $b \times p$ matrix, where p is the number of inputs, i.e., the size of current sources $I(t)$. $u(t)$ represents the nodal voltages of the network and $I(t)$ represents the current sources from the functional blocks. $\mathbf{C}$ and $\mathbf{A}$ are $n \times n$ matrices, where n is the number of nodes. Note that $\sigma(0)$ denotes the initial stress at $t = 0$. For each new simulation step, the stress from the previous simulation step is used as the initial condition, iteratively.

The above equations must be solved together. Linear network IR drop solver passes time-dependent current densities and power layout information to the FDTD EM solver. Once the voids are formed, IR drops in the power grid will change and the current at each time step will be different. The FDTD EM solver provides the IR drop solver with new resistance information, particularly, wires with voids. Since these two simulations are coupled together, the current and resistance of each mortal wire are dependent on each other. Note that $\mathbf{C}$ and $\mathbf{A}$ matrices that depend on wire structures are time-independent in the coupled equations.

In conclusion, *EMspice* is able to tell the resulted IR drop and EM failure hotspots at the target time. However, such analysis on a long target lifetime can be extremely time-consuming for very large power grid networks.

Chapter 3

EM immortality and lifetime constrained power grid wire sizing

3.1 EM immortality check and lifetime estimation for multi-segment wires

3.1.1 Steady-state EM-induced stress modeling

Steady-state EM-induced stress modeling helps find the immortality information of the interconnect wire quickly as no complex calculations are required. For these kinds of models, stress on the cathode at the steady state (σ_{steady}) is compared with critical stress (σ_{crit}). If σ_{steady} is lower than σ_{crit}, the wire is considered immortal. One of the well-known steady-state analysis methods is *Blech product* [10], but it is only suitable for a single (one-segment) wire. Recently, a voltage-based EM immortality analysis method for multi-segment interconnect structures has been proposed [50, 49]. In this method, an *EM*

voltage (V_E) which is proportional to stress at the ground node (σ_g) is calculated as

$$V_E = \frac{1}{2A} \sum_{k \neq g} a_k V_k \tag{3.1}$$

where V_k is the normal nodal voltage with respect to cathode node g at node k, a_k is the total area of branches connected to node k and A is the total area of the wire. With the voltage of node i (V_i), steady-state stress at that node (σ_i) is calculated as $\sigma_i = \beta(V_E - V_i)$, where $\beta = \frac{eZ}{\Omega}$, e is elementary charge, Z is effective charge number, and Ω is the atomic lattice volume. A *critical EM voltage* $V_{crit,EM}$ is defined by

$$V_{crit,EM} = \frac{1}{\beta}(\sigma_{crit} - \sigma_{init}) \tag{3.2}$$

where σ_{init} is the initial stress. In order to check whether the interconnect wire is immortal or not, we need to check the following condition

$$V_{crit,EM} > V_E - V_i \tag{3.3}$$

If this condition is met for all the nodes, EM failure will not happen. Since generally the cathode node has the lowest voltage within an interconnect wire, we may just check the cathode node instead of all the nodes, which means

$$V_{crit,EM} > V_E - V_{cat} \tag{3.4}$$

where V_{cat} is the voltage at the cathode.

The method can be illustrated using the following example. Fig. 3.1 shows a 3-terminal wire. In this wire, node 0 is treated as the ground node. The current densities of the two segments are j_a and j_b which may not be the same because they will be determined by the rest of the circuit. The EM stress equations become

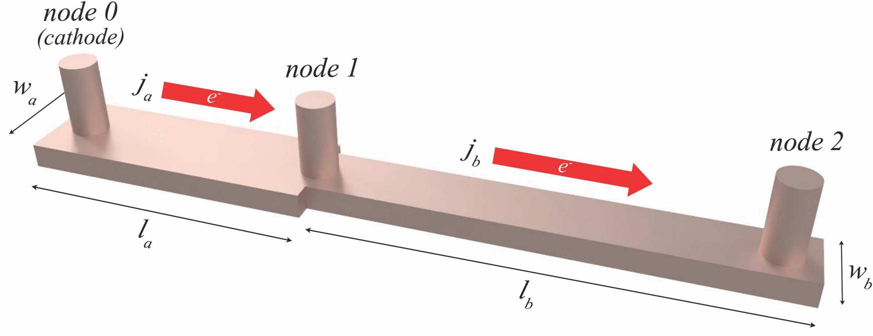


Figure 3.1: Interconnect example of a straight 3-terminal wire for EM analysis.

$$
\begin{aligned}
V_0 &= 0, & a_0 &= l_a w_a, & \sigma_0 &= \beta V_E \\
V_1 &= j_a l_a \rho, & a_1 &= l_a w_a + l_b w_b, & \sigma_1 &= \beta(V_E - V_1) \\
V_2 &= j_b l_b \rho + j_a l_a \rho, & a_2 &= l_b w_b, & \sigma_2 &= \beta(V_E - V_2)
\end{aligned} \tag{3.5}
$$

Then the *EM voltage* can be obtained easily

$$
V_E = \frac{a_0 V_0 + a_1 V_1 + a_2 V_2}{2A} = \frac{a_1 V_1 + a_2 V_2}{2A} \tag{3.6}
$$

where

$$
A = \frac{a_0 + a_1 + a_2}{2} \tag{3.7}
$$

We can compare V_E and $V_{crit,EM}$ to see if this wire is immortal.

3.1.2 Transient EM-induced stress estimation

Sometimes the steady-state methods are overly conservative, therefore more complete models of transient hydrostatic stress evolution are needed. In general, the failure process is divided into nucleation phase, incubation phase, and growth phase.

It is well-known that the nucleation phase was accurately modeled by Korhonen's equation [36], however, this PDE-based model is hard to solve directly using numerical methods and has very low efficiency for tree-based EM assessment. Recently a few numerical methods have been proposed such as the finite difference methods [20, 13] and analytical expressions-based approaches [15, 62]. In this work, an integral transformation method for straight multi-segment wires [62] is employed. Suppose we have n segments in a multi-segment wire as shown in Fig. 3.2. After discretizing Korhonen's equation, the stress can be expressed as

Figure 3.2: Example of a straight multi-segment wire.

$$\sigma(x,t) = \sum_{m=1}^{\infty} \frac{\psi_m(x)}{N(\lambda_m)} \bar{\sigma}(\lambda_m, t) \tag{3.8}$$

where the norm of eigenfunctions $N(\lambda_m)$ is

$$N(\lambda_m) = \int_{\chi=0}^{L} [\psi_m(\chi)]^2 d\chi \tag{3.9}$$

and $\bar{\sigma}(\lambda_m, t)$ is transformed solution of stress which is

$$\begin{aligned} \bar{\sigma}(\lambda_m, t) =& \bar{F}(\lambda_m) e^{-\kappa\lambda_m^2 t} + \frac{1}{\lambda_m^2}(1 - e^{\kappa\lambda_m^2 t}) \\ & \cdot \sum_{k=1}^{n} \Gamma_k \cdot \left(cos\frac{x_{k-1}}{L} m\pi - cos\frac{x_k}{L} m\pi \right) \end{aligned} \tag{3.10}$$

where F is

$$\bar{F}(\lambda_m) = \int_{\chi=0}^{L} \psi_m(\chi) \cdot \sigma_0(\chi) d\chi \tag{3.11}$$

λ_m and $\psi(x)$ are eigenvalues and eigenfunctions which are the solutions of the Sturm-Liouville problem corresponding to the diffusion Eq. (2.4) and the boundary conditions

$$\lambda_m = \frac{m\pi}{L}, \;\; \psi_m(x) = cos\frac{x}{L}m\pi \tag{3.12}$$

where $m = 1, 2, \cdots, \infty$. Γ_k is

$$\Gamma_k = \frac{eZ\rho}{\Omega} j_k, k = 0, 1, \ldots, n \tag{3.13}$$

With Eq. (3.8), given critical stress σ_{crit}, the nucleation time t_{nuc} can be obtained quickly by using nonlinear equation solving methods such as Newton's method or bisection method.

We remark that when the compressive stress at the anode continues to be built up, hillocks or extrusion may be formed, which will lead to a resistance decrease [60] and can potentially cause short-circuit failure. But the void nucleation is still the dominant EM failure effect [32]. In this work, we only focus on the void-induced EM failure.

After the void is nucleated, the incubation phase starts. In this phase, the resistance of the interconnect remains almost unchanged since the cross-section of the via is not covered by the void and the current can still flow through the copper.

In power grid networks, the interconnect trees are generally multi-segment wires. All segments connected with the void can contribute to the void growth since electron wind at each segment can accelerate or slow down the void growth based on their directions. In this phase, void growth rate v_d is estimated to be [51]

$$v_d = \frac{D_a eZ\rho}{kTW_m} \sum_i j_i W_i \tag{3.14}$$

where j_i and W_i are the current density and width of the ith segment, respectively. W_m is the width of the main segment where the void is formed.

After the incubation phase, the void fully covers the via, initiating the growth phase. In this phase, the resistance starts increasing. It is important to note that early failure and late failure have different failure mechanisms.

For early failure, the wire fails once the void covers the via, which means that the wire fails at the end of the incubation phase and there is no growth phase ($t_{growth} = 0$). Hence, the failure time is the sum of t_{nuc} and t_{inc}.

For late failure, after the void size reaches the critical size, there will be no open circuit since the current can still flow through the barrier layer. In this case, the void growth will lead to resistance increase. When the resistance increases to the critical level, the interconnect wire is considered to be failed. The time period between the void covering the via and wire failure is called growth time. The growth time for late failure is

$$t_{growth} = \frac{\Delta r(t)}{v_d \left[\frac{\rho_{Ta}}{h_{Ta}\left(2H + W\right)} - \frac{\rho_{Cu}}{HW} \right]} \tag{3.15}$$

where ρ_{Ta} and ρ_{Cu} are the resistivities of tantalum (the barrier liner material) and copper, respectively, W is the line width, H is the copper thickness, and h_{Ta} is the liner layer thickness.

However, the void may saturate before reaching the *critical void length*. The saturation length is expressed in [33] as

$$L_{ss} = L_{line} \times \left[\frac{\sigma_T}{B} + \frac{eZ\rho jL}{2B\Omega} \right] \tag{3.16}$$

where L_{ss} is the void saturated length, L_{line} is the total length of the wire and σ_T is thermal stress. Void growth may stop before the calculated t_{growth} because the void is saturated. In this case, we can treat the wire as immortal or its lifetime is larger than the target lifetime.

3.2 EM immortality constrained optimization for multi-segment interconnects

Study and experimental data show that the current-induced stress developed in the individual segments within an interconnect tree is not independent. In other words, if we just look at the current density for each segment individually, it may appear as if all wire segments are immortal, but the whole interconnect tree could still be mortal. The reason is that the stress in one segment of an interconnect tree depends on other segments, which are not independent. As discussed before, this issue has been resolved by the fast EM immortality check method for general multi-segment interconnect wires.

In this section, we present the EM immortality constrained power grid wire sizing optimization method considering multi-segment interconnect wires. We notice that the new EM constraint will ensure that all the wires are EM immortal, so we call this method EM immortality constrained power grid optimization.

3.2.1 Problem formulation

Let $G = \{N, B\}$ be a power grid network with n nodes $N = \{1, ..., n\}$ and b branches $B = \{1, ..., b\}$. Each branch i in B connects two nodes i_1 and i_2 with current flowing from i_1 to i_2. l_i and w_i are the length and width of branch i, respectively. ρ is the

sheet resistivity. The resistance r_i of branch i is

$$r_i = \frac{V_{i_1} - V_{i_2}}{I_i} = \rho \frac{l_i}{w_i} \tag{3.17}$$

Objective function

We can express the total routing area of a power grid network in terms of voltages, currents and lengths of branches as follows:

$$f(V, I) = \sum_{i \in B} l_i w_i = \sum_{i \in B} \frac{\rho I_i l_i^2}{V_{i1} - V_{i2}} \tag{3.18}$$

We notice that the objective function is linear for branch current variables **I** and nonlinear for node voltage variables **V**.

The EM immortality constrained optimization in Eq. (3.18) does not depend on temperature. In other words, the immortality condition of a wire is temperature-independent. However, if the wire fails, the nucleation time will depend on the temperature, which will be addressed by the lifetime constrained optimization shown later.

Constraints

The constraints that need to be satisfied for a reliable, working power grid network are shown as follows.

Voltage IR drop constraints In order to ensure proper logic operation, the IR drop from the power pads to the nodes should be restricted. For each node, we must specify a threshold voltage

$$V_j > V_{min} \text{ for power network} \tag{3.19}$$

In the modern high-performance processor design, usually 10% or so of the overall wiring resources on all wiring levels are dedicated to power delivery.

Minimum width constraints The widths of the power stripes are technologically limited to the minimum width allowed for the layer where the segment lies. Thus, we have

$$w_i = \rho \frac{l_i I_i}{V_{i1} - V_{i2}} \geq w_{i,min} \tag{3.20}$$

EM constraints for multi-segment interconnects As described before, for a multi-segment interconnect m, the EM constraint should be satisfied

$$V_{crit,EM} > V_{E,m} - V_{cat,m} \tag{3.21}$$

where $V_{E,m}$, computed by Eq. (3.1), is the *EM voltage* for the mth interconnect tree and $V_{cat,m}$ is the cathode nodal voltage of that tree. Unlike previous methods whose branch currents are monitored and used as constants, in this new method, voltages are used as constraints. Only the cathode node voltage for a whole interconnect tree needs to be monitored and no other complex calculations are required.

Equal width constraints For typical chip layout designs, certain tree branches should have the same width. The constraint can be written as $w_i = w_k$. In terms of nodal voltages and branch currents, we have

$$\frac{V_{i1} - V_{i2}}{l_i I_i} = \frac{V_{k1} - V_{k2}}{l_k I_k} \tag{3.22}$$

Kirchhoff's current law (KCL) For each node j, we have

$$\sum_{k \in B(j)} I_k = 0 \tag{3.23}$$

where $B(j)$ is the set of branches incident on node j.

The power grid optimization aims to minimize objective function (3.18) subjected to constraints (3.19)-(3.23). It will be referred as problem P. Problem P is a constrained nonlinear optimization problem.

3.2.2 Relaxed two-step sequence of linear programming solution

In the aforementioned constrained nonlinear optimization problem, we notice that the EM constraint (3.21) is still linear in terms of nodal voltage. As a result, we can follow the relaxed two-phase iterative optimization process [16, 57] and apply the sequence of linear programming technique to solve the relaxed problem in each phase. Specifically, we have two phases: the voltage solving phase (P-V phase) and the current solving phase (P-I phase).

P-V optimization phase

In this phase, we assume that all branch currents are fixed, then the objective function can be rewritten as

$$f(V) = \sum_{i \in B} \frac{\alpha_i}{V_{i1} - V_{i2}} \tag{3.24}$$

where $\alpha_i = \rho I_i l_i^2$, subject to constraints (3.19), (3.21) and (3.22). We further restrict the changes of nodal voltages such that their current directions do not change during the

optimization process:

$$\frac{V_{i1} - V_{i2}}{I_i} \geq 0 \tag{3.25}$$

Problem P-V is nonlinear, however, it can be converted to a sequence of linear programming problem [57]. By taking the first-order Taylor's expansion of Eq. (3.24) around the initial solution V^0, the linearized objective function can be written as

$$g(V) = \sum_{i \in B} \frac{2\alpha_i}{V_{i1}^0 - V_{i2}^0} - \sum_{i \in B} \frac{\alpha_i}{\left(V_{i1}^0 - V_{i2}^0\right)^2} (V_{i1} - V_{i2}) \tag{3.26}$$

Besides, an additional constraint will be added

$$\xi sign(I_i) \left(V_{i1}^0 - V_{i2}^0\right) \leq sign((I_i) (V_{i1} - V_{i2}) \tag{3.27}$$

where $\xi \in (0, 1)$ is a restriction factor, which will be selected by some trials and experience and $sign(x)$ is the sign function.

P-I optimization phase

In this phase, we assume that all nodal voltages are fixed, so the objective function becomes

$$f(I) = \sum_{i \in B} \beta_i I_i \tag{3.28}$$

where $\beta_i = \dfrac{\rho l_i^2}{V_{i1} - V_{i2}}$, subject to constraints (3.20), (3.22) and (3.23). Similarly, we restrict the changes of current directions during the optimization process:

$$\frac{I_i}{V_{i1} - V_{i2}} \geq 0 \tag{3.29}$$

As can be seen, problem P-I is a linear programming problem.

3.2.3 EM immortality constrained power grid optimization

The EM immortality constrained power grid optimization starts with an initial feasible solution. We iteratively solve P-V and P-I. The procedure for solving the problem in P-V phase can be transformed to the problem of choosing ξ and solving a linear programming problem, and then repeating this process until the optimal solution is found. We summarize the entire EM immortality constrained power grid optimization procedure as Algorithm 1.

Algorithm 1 EM immortality constrained power grid wire sizing algorithm

Input: Spice netlist G_I containing a power grid network.

Output: Optimized power grid network parameters.

1: /*Problem Setup*/

2: $k := 0$.

3: Compute the initial V^k, I^k from G_I.

4: **repeat**

5: /*P-V Phase*/

6: Construct constraints (3.21), (3.22), (3.25) and (3.27) with I^k.

7: $m := 1$.

8: Compute $V_m^k := \arg\min g\left(V^k\right)$ subject to constraints (3.19), (3.21), (3.22), (3.25) and (3.27).

9: **while** $f\left(V_m^k\right) > f\left(V_{m-1}^k\right)$ **do**

10: Determine the search direction $d_m := V_{m-1}^k - V_m^k$.

11: Choose step size α for line search.

12: $V_{m+1}^k := V_m^k + \alpha d_m$.

13: $m := m + 1$.

14: **end while**

15: $V^{k+1} := V_m^k$.

16: /*P-I Phase*/

17: Construct constraints (3.20), (3.22) and (3.29) with V^{k+1}.

18: Compute $I^{k+1} := \arg\min f\left(I^k\right)$ subject to (3.20), (3.22), (3.23), and (3.29) constraints.

19: $k := k + 1$.

20: **until** $\left|f\left(V^k, I^k\right) - f\left(V^{k-1}, I^{k-1}\right)\right| < \varepsilon$

In practice, only a few linear programmings are needed to reach the optimum solution. Thus the time complexity of our method is proportional to the complexity of linear programming. It is known that linear programs can be solved in polynomial time using the interior point method [57, 8], which will be represented as $f_{SLP}(n, b)$ from now where n and b are the number of nodes and number of branches respectively in the given power grid network.

We remark that in Algorithm 1, the nodal voltages and branch currents will not change dramatically from one iteration to another for a similar value of the objective function (3.18). The reason is that the branch voltages and node voltages are uniquely related in the power and ground networks where the reference voltages are given. If the voltage difference or branch voltage for a wire will not change dramatically from one iteration to another during the optimization, which typically is the case for the late stages of P-V optimization, then the nodal voltages will not change dramatically as well. Moreover, the nodal voltages will be uniquely determined by those branches and reference voltages. This is also

true for the P-I phase as the branch current solution needs to satisfy Kirchhoff's current law first. On top of this, if the branch voltages do not change too much in the P-V phase based on the previous analysis, the branch currents will not change significantly as well for a similar value of the objective function.

3.3 EM immortality constrained power grid wire sizing with reservoir insertion

As mentioned earlier, EM immortality requirement for all the interconnect trees can be overconservative. This happens when the initial power grids are designed so that we may not be able to find a solution from the wire sizing method in Algorithm 1. To mitigate the issue, we present a novel reservoir insertion-based method.

Reservoir is a zero-current metal segment added to the cathode of a metal wire so that the wire's lifetime can be increased. Fig. 3.3 shows a two-segment wire in which the left segment is a reservoir. Adding a reservoir branch will reduce the steady state tensile stress of the cathode node, which may make the mortal wire immortal. In other words, it can significantly increase the lifetime of the wire. The area of the reservoir will determine the final steady state stress at the cathode node. Fig. 3.4 shows how the different widths and lengths of the reservoir will affect the TTF of the reservoir-enabled wires. As we can see, increasing the length or width of the reservoir will benefit the lifetime. Doubling the length increases nucleation time by around 13.7% while doubling the width increases nucleation time by 89.5% [51].

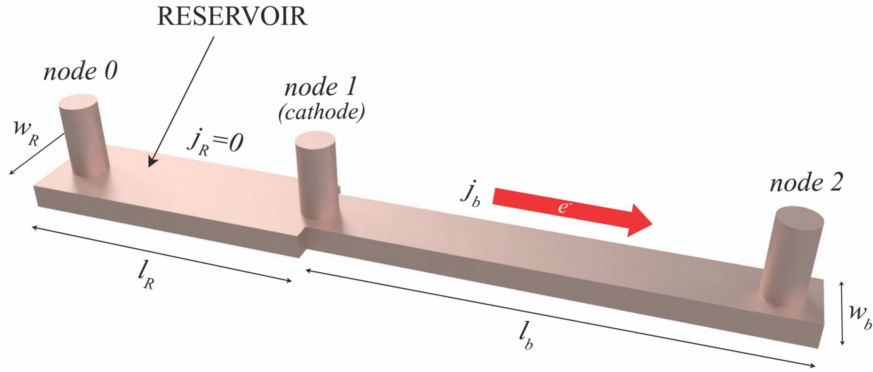


Figure 3.3: The 3D view of a two-segment wire with a reservoir.

With the impact on TTF shown in Fig. 3.4, we can adjust the width or length of a reservoir to change the lifetime of the wire. However, there will be an opposite impact on the lifetime when a zero-current wire is added to the anode.

Fig. 3.5 shows the placement of reservoirs in a practical power grid layout where the shaded yellow branches in the power grids are reservoirs. If the cathode is at the terminal of a multi-branch wire, then the reservoir can be placed either at the top, bottom, left, or right of the wire. In general, reservoirs should be placed as close to the cathode as possible to have the maximum effect. But in a practical layout, the reservoir placement is subject to routing and area constraints. After the location of the reservoir is determined, we need to decide its size and shape. As we know, the width of the reservoir has a higher impact on the lifetime, to be more specific, on the nucleation time. The wider the width, the longer the lifetime as it leads to slower stress growth. As a result, width is a preferred parameter over length to improve the EM lifetime. Take Fig. 3.3 as an example, assume node 1 is the

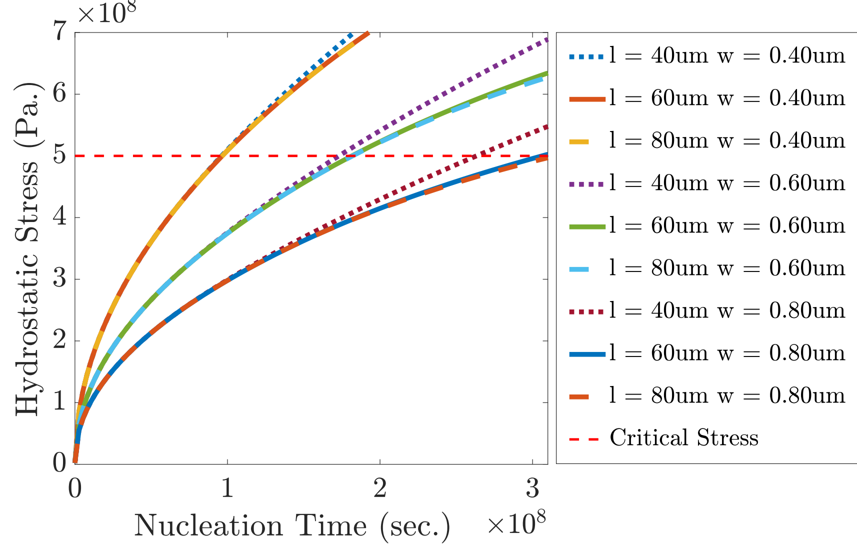


Figure 3.4: The impact of length and width of the reservoir on the TTF of a wire.

ground (cathode) node, then j_R should be 0. We can obtain the following equations:

$$\begin{aligned} V_E &= \frac{a_2 V_2}{2A} = \frac{(l_b w_b) \cdot V_2}{(l_b w_b) + (l_b w_b)} \\ V_{E,new} &= \frac{a_2 V_2}{2A_{new}} = \frac{(l_b w_b) \cdot V_2}{(l_R w_R) + (l_R w_R + l_b w_b) + (l_b w_b)} \end{aligned} \tag{3.30}$$

where A and A_{new} are the areas of the wire before and after reservoir insertion respectively.

According to Eq. (3.4), to make the interconnect immortal, we should have

$$V_{crit,EM} > V_{E,new} - V_{cat,m} \tag{3.31}$$

Here $V_{cat,m}$ will be zero for ground networks, but will not be zero for power networks for the mth interconnect. Then A_{new} can be calculated easily by making the inequality equal

$$A_{new} = \frac{A \cdot V_E}{V_{crit,EM} + V_{cat,m}} \tag{3.32}$$

and $A_{reservoir}$ can also be obtained

$$A_{reservoir} = A_{new} - A \tag{3.33}$$

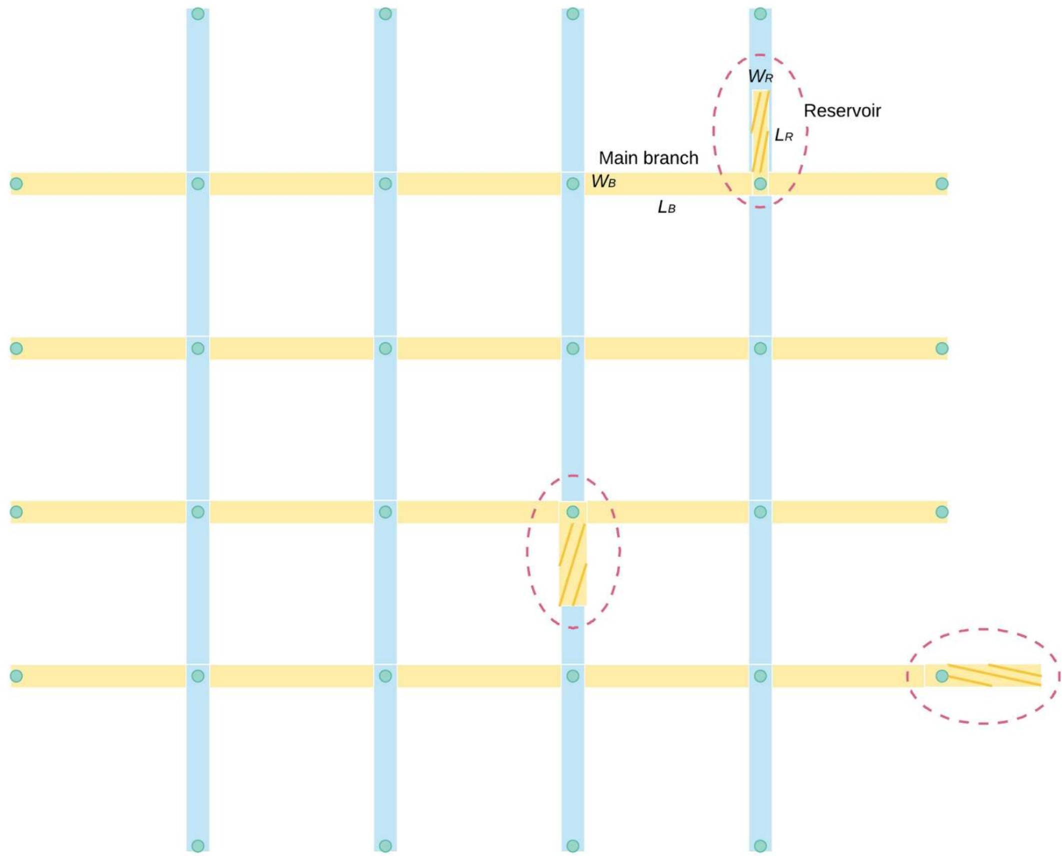


Figure 3.5: Inserted reservoirs in the power grid network.

$A_{reservoir}$ is actually the minimum area added to the cathode node to make the mortal wire immortal. For power grid network optimization, adding reservoir branches helps increase the lifetime of the wire greatly, at the cost of some chip area. In other words, for the wires that are hard to optimize with the EM immortality constrained optimization, we can trade some chip area for a longer power wire lifetime. The modified wire sizing algorithm is summarized as Algorithm 2.

Algorithm 2 EM immortality constrained power grid wire sizing algorithm with reservoir insertion

Input: Spice netlist G_I containing a power grid network.

Output: Optimized power grid network parameters.

1: **repeat**

2: Perform *Algorithm 1* using G_I.

3: **if** fail due to the EM constraints **then**

4: Find all the mortal wires.

5: Insert reservoir segments (also properly sized) at or around the cathodes of these wires so that all of those wires become immortal.

6: **else**

7: Break.

8: **end if**

9: **until** optimize successfully

When compared to Algorithm 1, Algorithm 2 needs to do EM condition check and reservoir insertion, whose computing cost is quite small and thus can be ignored. As a result, Algorithm 2 has the same time complexity as Algorithm 1: $f_{SLP}(n, b)$.

Note, once a reservoir is inserted into a power grid network, it does not participate in sizing, but just provides a smaller *EM voltage*. The reason being, the reservoir is a zero-current branch, and in the objective function we represent branch area by branch current in the numerator.

Adding reservoirs may be subject to some design rules in the practical layout. We first try to add the reservoir branch along the preferred routing direction, which is

perpendicular to the current flow direction of the main branch. The maximum allowable length for the reservoir branch is the grid distance minus the minimum space requirement as shown in Fig. 3.6. Otherwise, we can add the reservoir branch in the non-preferred direction if the design rule is still allowed. If neither is possible, then the EM lifetime constrained power grid optimization method will be applied.

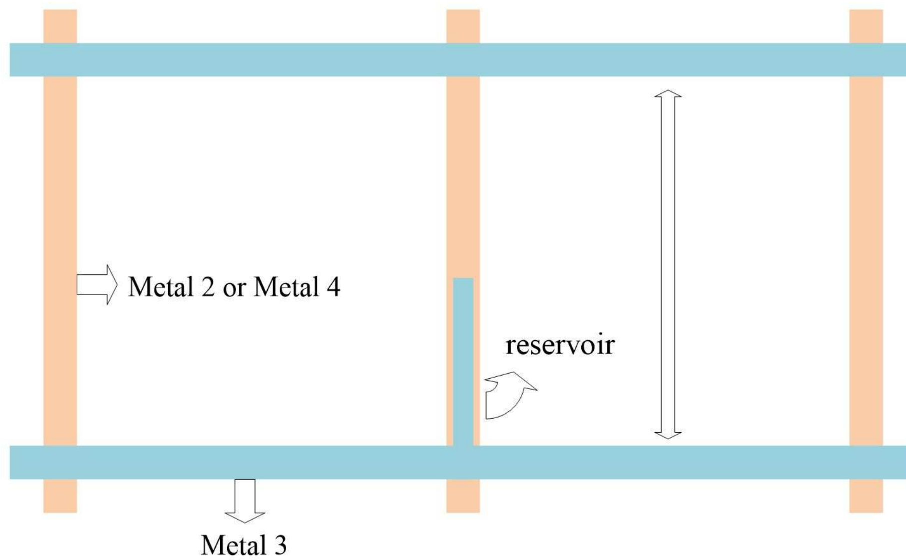


Figure 3.6: DRC check for reservoir segment placement.

3.4 EM lifetime constrained power grid optimization

In the previous sections, we discussed the power grid sizing optimization ensuring none of the interconnect trees fail based on the voltage-based EM immortality check. However, such EM constraint may be conservative because some wires can have EM failures as long as the power grid networks can still work (its IR drop is still less than the given threshold) at the target lifetime.

3.4.1 EM lifetime constrained power grid optimization flow

In this section, we present a new EM lifetime constrained power grids wire sizing optimization method in which some segments of multi-segment interconnect wires will be allowed to fail or to age. The impacts of these segments in terms of resistance change or even wire openings will be explicitly considered and modeled. Such aging-aware EM optimization essentially takes the EM aging-induced impacts or guard bands into account so that the designed power grid networks can still work in the target lifetime. In this work, we only consider void formation, which is the dominant EM failure effect and will lead to an increase in resistance. On the other hand, hillocks or extrusion will cause resistance decrease or even short-circuit, i.e., $\Delta R < 0$. In terms of voltage and current, $\Delta R < 0$ means a decrease in voltage or IR drop, or an increase in current, or both. As a result, hillocks can have positive (reduced IR drop) or negative impacts (increased current density) on the power grid network.

The new optimization flow is shown in Fig. 3.7. In this new flow, we first check whether a given power grid network can be optimized using the EM immortality power grid optimization given in Algorithm 1. If the optimization does not succeed, it means the EM condition may be over-constrained (perhaps the over-constraints are due to IR drop constraints instead of EM constraints, in this case, topology changes will be needed and it falls out of the scope of this paper). After this step, the lifetime of all the interconnect trees will be computed based on the EM lifetime estimation method discussed in Algorithm 3.

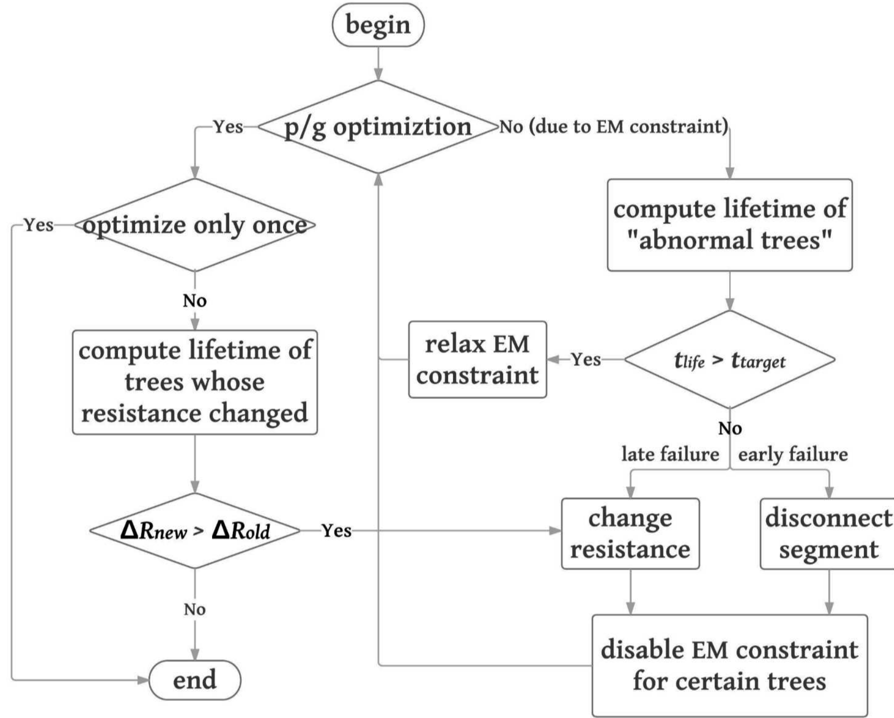


Figure 3.7: Flowchart of the EM lifetime constrained power grid optimization process.

We have several scenarios to discuss before we perform the optimization again. Let's define $t_{life,m}$ as the lifetime of the mth interconnect tree and t_{target} as the target lifetime.

1. **If $V_{E,m} - V_{cat,m} > V_{crit,EM}$ and $t_{life,m} < t_{target}$:**

   The mth interconnect wire will be marked as a *failed wire*. Then we have the following changes for the wire before the next round of optimization.

   If it is an early failure case, the cathode node of the wire segment connected by the failed via will be disconnected, which is called *wire disconnection*. The failure cases will depend on the current directions around the cathode node. Also, the disconnection

will depend on whether the void growth can eventually reach the critical void size or not as discussed in Chapter 3.1.

If it is a late failure case, the wire segment associated with the cathode node will have a *resistance change*. The specific resistance change for each failed segment will be calculated based on the target lifetime using our EM lifetime estimation method.

If it is a late failure case, the wire segment associated with the cathode node will have a *resistance change*. The specific resistance change for each failed segment will be calculated based on the target lifetime using our EM lifetime estimation method.

2. **If $V_{E,m} - V_{cat,m} > V_{crit,EM}$ and $t_{life,m} > t_{target}$:**

   The lifetime of interconnect wire still meets the target lifetime even though it will have void nucleation and resistance change. This also includes the case in which void growth saturates before its size reaches the critical void size. The wire still works since the current can flow through the barrier layer.

   In this case, we use the existing $V_{E,m} - V_{cat,m}$ value as the new EM constraint (defined as $V_{E,m,next} - V_{cat,m,next}$) for this wire only ($m$th wire): $V_{E,m} - V_{cat,m} < V_{E,m,next} - V_{cat,m,next}$. This is called *constraint relaxation*. The rationale behind this is that we expect the EM status of this wire to become worse during the next optimization so its lifetime will not change too much and still meet the given lifetime after the follow-up optimizations.

After *resistance change*, or *wire disconnection*, or *constraint relaxation*, a new round of SLP programming optimization, which is similar to Algorithm 1, is carried out.

When the power grid optimization is successfully finished, one has to check if all the wires, which are marked as failed ones, meet their failure conditions. For instance, if a failed wire (its failed segment) has a 20% resistance change at the target 10-year lifetime before the optimization, however, after the optimization, it has 30% resistance change at 10 years, then the final IR drop condition may not meet at 10 years. As a result, more iterations need to be carried out until such process is converged.

Another way to avoid such iterations is to set up some guard bands. Specifically, we can increase the target lifetime from 10 years to 15 years when computing the resistance change for all the failed wires. After optimization, we check whether all the failure conditions are satisfied against the real 10-year target. In this way, we may just need one iteration at the cost of more routing areas used for the power grid networks.

The time complexity of the new EM lifetime constrained optimization flow shown in Fig. 3.7 is $O(l * f_{SLP}(n, b)) + O(b)$, where the l is the number of iterations of Algorithm 1 and $O(b)$ comes from Algorithm 3 discussed below. Since $f_{SLP}(n, b)$ is nonlinear with respect to b, the total time complexity can be further reduced to $O(l * f_{SLP}(n, b))$.

3.4.2 Transient EM analysis for a multi-segment interconnect

One important aspect of the above lifetime constrained power grid network optimization is to calculate the lifetime of a given wire and its electrical conditions. Another process we need is to compute the resistance change for a given target lifetime. Based on the EM models and fast assessment techniques presented in Chapter 3.1, the two functions can be described by the following transient EM analysis algorithm: Algorithm 3. In this

algorithm, if the increased resistance of the nucleated branch exceeds 10%, the interconnect tree is marked as failed.

Algorithm 3 Lifetime or resistance change analysis for an interconnect tree

Input: A given interconnect wire and its electrical conditions, target lifetime t_{target}.

Output: Lifetime of the wire t_{life} and resistance change ΔR at t_{target}.

1: Compute the initial current density and the corresponding EM driving force.
2: /*Nucleation Phase*/
3: Compute the transient hydrostatic stresses $\sigma_{x,t}$ inside the interconnect tree.
4: **if** $\sigma_{max,t} > \sigma_{crit}$ **then**
5:     Solve the nucleation time t_{nuc} using the bisection method.
6:     /*Incubation Phase*/
7:     **if** $L_{ss} < \Delta L_{crit}$ **then**
8:         $t_{inc} := \infty$ and $\Delta R := 0$.
9:     **else**
10:         Compute t_{inc} using Eq. (2.5).
11:         /*Growth Phase*/
12:         **if** early failure mode is true **then**
13:             $t_{growth} := 0$ and $\Delta R := \infty$.
14:         **else**
15:             Compute t_{growth} using Eq. (3.15) assuming $\Delta R = 0.1 * R$ or compute the wire resistance change ΔR so that $t_{nuc} + t_{inc} + t_{growth} = t_{target}$.
16:         **end if**
17:     **end if**

18: $t_{life} := t_{nuc} + t_{inc} + t_{growth}$.

19: **else**

20: $t_{life} := \infty$ and $\Delta R := 0$.

21: **end if**

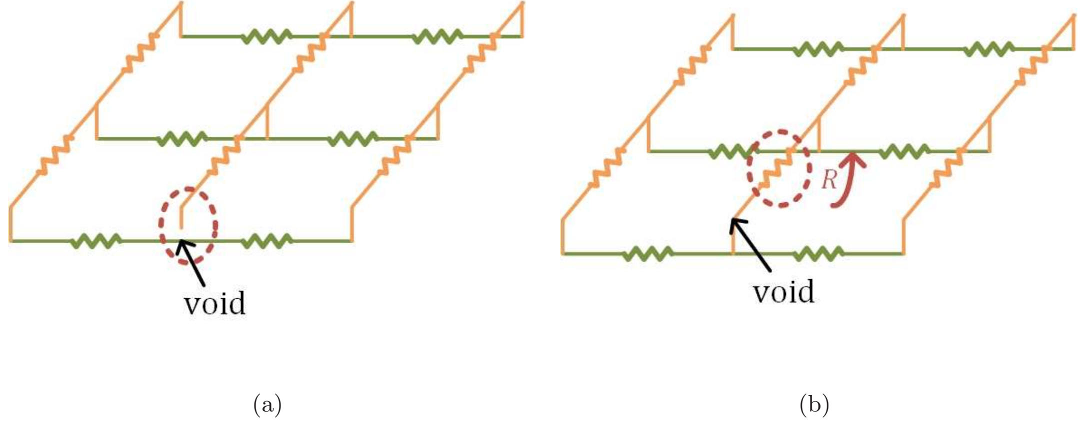
In Algorithm 3, to compute the lifetime t_{life} of a given wire, we need to make sure that the wire is mortal and Eq. (3.4) is not satisfied. If the target lifetime t_{target} is given, then Algorithm 3 will return the resistance change ΔR at the target lifetime.

For those mortal wires, we start with time $t = 1000$ years and use the bisection method to find t_{nuc}. The transient hydrostatic stress will be computed by Eq. (3.8) at the given time. Once the stress of one segment hits the critical stress, the wire is deemed as nucleated.

First, we discuss the case that the wire is void incubation phase immortal, that is, the void saturated length is less than the critical length. Then incubation time (eventually the lifetime) becomes infinite and the resistance remains unchanged.

Otherwise, the failure mode of the wire should be determined by looking at the current direction in the cathode node based on the patterns in Fig. 2.3(a) and Fig. 2.3(b).

If the wire is in early failure mode, then the wire will become an open circuit. In this case, the whole interconnect tree will be disconnected from another interconnect wire as shown in Fig. 3.8(a). For the wire in late failure mode, we have another solution. The wire resistance change will be incurred and the growth time will be computed when the resistance change reaches 10% as shown in Fig. 3.8(b). If the target lifetime is given, then the wire resistance changes by ΔR at the target lifetime.



(a) (b)

Figure 3.8: The electrical impact of different failure mechanisms on the interconnect wires: (a) early failure mode; (b) late failure mode.

In Algorithm 3, it is clear that finding out the void nucleation time t_{nuc} is the most time-consuming part. Specifically, the time complexity of calculating hydrostatic stress for one multi-segment wire can be expressed as $O(s_i)$, where s_i is the number of segments of that wire. Once we obtained the stress, the cost of using the bisection method to find t_{nuc} is very small and can be ignored. Overall, the time complexity of Algorithm 3, which is to find the nucleation time for wire i, is $f_{Alg3,i} = O(s_i)$. If in the worst case, all the wires are calculated, then total time complexity becomes $O(b) = \sum_i^b f_{Alg3,i}$, where b is the number of wire segments of the whole power grid network.

3.5 Experimental results and discussions

3.5.1 Experiment setup

The presented EM-aware and lifetime constrained power grid optimization is implemented in C/C++. We carried out the optimization on a workstation with 2 Intel Xeon E5-2698 which has 128G memory. Both IBM power grid networks [43] and self-generated

networks are used to test our work. From the annotated Spice format IBM benchmarks, current source, voltage source, and resistance values can be obtained easily. Node location and layer information are also provided, therefore, the geometric structure of a certain benchmark is known. Since there is no wire length and width information, we made some assumptions including wire length and layer height for the experiments. Within the IBM power grid benchmarks, there are both power networks and ground networks. Only power networks are used to test our method and their current source values are scaled to ensure that the initial voltage drop of each node is small enough. We assume that the material of the multi-segment interconnect wire is Cu and the barrier layer is Ta. The maximum allowable IR drop is assumed to be $10\%V_{dd}$ and the minimum allowable width is $0.1\mu m$. Some important parameters used in our experimental setup are listed in Table 3.1.

Table 3.1: Parameters used in the optimization process.

parameter	value	parameter	value
σ_{crit}	500MPa	Ω	$1.182 \times 10^{-29}\mathrm{m}^3$
V_{crit}	$3.69 \times 10^{-3}\mathrm{V}$	E_a	0.8eV
T	323K	D_0	$5.55 \times 10^{-8}\mathrm{m}^2/\mathrm{s}$
B	140GPa	ρ_{Cu}	$1.9 \times 10^{-8}\Omega \cdot \mathrm{m}$
Z	10	ρ_{Ta}	$1.35 \times 10^{-7}\Omega \cdot \mathrm{m}$

3.5.2 EM immortality constrained power grid optimization results

Table 3.2 compares the results of the new optimization method using immortal EM constraint with the results of the existing power grid optimization method using the current density only EM constraint [58, 57]. Although many EM models have been proposed, including the new physics-based models [55], the current density-based EM models based

on the Black equation and Blech's product [9, 10] are still widely used in today's industry even though they are subject to growing criticism. For the full-chip EM analysis, most EDA tools only check the current density to see if it meets the requirements provided by the manufacturer. Therefore, comparing our method against the SLP method based on the current-based model is still meaningful. The new EM immortality model [50, 49] used in this SLP algorithm can be viewed as the extension of Blech's product to general multi-segment interconnects. As a result, essentially we compare and evaluate the impact of the EM immortality model against the Blech's product-based traditional EM model.

Table 3.2: EM immortality constrained power grid optimization results for IBM power grid networks.

circuit	# node	# bch	area (mm^2)	VB-EM constraint		CD-EM constraint		
				CPU time (min)	area reduced (%)	CPU time (min)	area reduced (%)	$t_{life-min}$ (yr)
ibmpg1	11572	5580	158.43	0.06	35.66	0.06	72.29	13.57
ibmpg2	61797	61143	60.38	3.13	77.55	1.65	91.35	8.16
ibmpg3	407279	399201	697.71	131.00	22.60	28.57	57.98	9.64
ibmpg4	474836	384709	210.44	186.20	18.42	112.38	29.70	7.61

In our optimization processes, we assume that all the branches have the same width on one tree but different trees can have different widths. For a fair comparison, we only allow the difference in the EM constraints. In Table 3.2, column 1 to 4 list the power grid network benchmarks (*circuit*), the number of nodes in the power grid network (*# node*), the number of branches (*# bch*), and the original area (*area (mm^2)*). Column 5 and 7 show the CPU time of the two methods in minutes (*CPU time (min)*). Column 6 and 8 report the reduced chip area ratio (*area reduced (%)*) with respect to the original area for the two methods with voltage-based EM constraint (*VB-EM constraint*) and current density-

based EM constraint (*CD-EM constraint*), respectively. Column 9 presents the minimum lifetime in years ($t_{life-min}$ *(yr)*) from current density-based EM power grid optimization as we may have mortal wires after the optimization. Usually, the optimization process can be completed within 3 iterations.

As we can see, for the *ibmpg2* example, which has 61797 nodes, 120 voltage sources and 18963 current sources, the original area is 60.38mm^2. After 2 iterations, the area can be reduced to 13.55mm^2, which means 77.55% area has been reduced. Although the previous work [57] has about 91.35% reduction, better performance in terms of area reduction, our detailed analysis shows that there exist many mortal wires. For instance, one wire has a lifetime of 7.8 years. In contrast, the new work ensures that none of the wires fail.

It can be observed that the new method typically leads to less area reduction compared to the current density-based method. But we want to emphasize that the two methods are not equal as the existing current density method [58] was based on Black's equation, which cannot ensure EM immortality. Table 3.2 basically shows that the presented method can guarantee the immorality of the whole power grid network, while the existing method cannot and some of the optimized results can even lead to the violation of 10-year lifetime constraint (such as the *ibmpg2, ibmpg3, ibmpg4*) even though they may have better area reduction ratio. Essentially the current-based EM optimization trades lifetime for better area reduction.

As for CPU time, since all interconnect trees need to be calculated at least once in the new power grid optimization method, it takes some time to build the new EM constraint, as a result, the new method is a bit slower than the *CD-EM constraint* method.

Furthermore, interconnect trees with more branches usually need more time to get the EM voltage.

3.5.3 EM immortality constrained power grid optimization with reservoir insertion results.

Next, we show the results from the reservoir insertion and EM immortality constrained SLP programming for power grid optimization. Instead of directly using the original IBM power grid networks, we made some modifications to them. We mainly increased the lengths of wires while keeping the original mesh structures. Additionally, a few branches of the benchmarks have been added or deleted. The goal is to make the power grid more vulnerable to EM failure.

The only exception is *ibmpg4a*, which is the same as *ibmpg4*. In this case, we show that if a power grid network can be optimized successfully without reservoir insertion, then performing Algorithm 2 is equivalent to performing Algorithm 1.

In the optimization process, we assume all the reservoirs are added in two directions, e.g., on top of a horizontal interconnect tree and on the left of a vertical interconnect tree. If one wire is determined to be mortal, we first add a reservoir of calculated size at the cathode node to make the wire immortal and then perform the two-phase optimization process. The inserted reservoir and the interconnect tree should be in the same layer. The length of the reservoir will be limited to be less than its perpendicular branch length.

Table 3.3 shows the results of the new optimization method with the reservoir insertion feature. In this table, column 2 and 3 show the number of total trees (*# tree*) and the number of trees that may fail (*# nucleated wires*). Column 5 tells if the first

optimization (*w/o reservoir insertion*) can be finished successfully. Column 6 and 7 give the CPU time and the reduced chip area ratio with respect to the original area when the reservoir insertion option is turned on.

Table 3.3: EM immortality constrained power grid optimization with reservoir insertion results for IBM power grid networks.

circuit	# tree	# nuc-wires	area (mm^2)	w/o reservoir	w reservoir	
				finish successfully?	CPU time (min)	area reduced (%)
ibmpg1a	689	249	633.56	no	0.06	21.25
ibmpg2a	462	91	120.65	no	3.15	28.64
ibmpg3a	7388	329	1032.55	no	131.85	15.22
ibmpg4a	9458	0	210.44	yes	186.20	18.42

As we can see, the CPU time results are similar to those with *VB-EM constraint* method in Table 3.2, which means that reservoir insertion does not have much computing overhead. In Table 3.3, the power grid networks are quite different from those in Table 3.2 because of the EM concern. So the area reduction results of the two networks are different even using the same SLP algorithm.

From the results, we can see that by adding reservoir branches, the mortal power grid networks can be made immortal without consuming much extra running time, and the area can still be optimized successfully using the sequence of linear programming. In every iteration, nodal voltages and branch currents are changed. We remark that cathode node voltage may change, but the cathode node typically does not change. After adding a reservoir onto an interconnect, a new EM voltage $V_{E,new}$ in Eq. (3.30) will be used for power grid optimization. Then the area of the resulting interconnect will be optimized (reduced or no change) while the EM immortality constraint can still hold after the optimization.

3.5.4 EM lifetime constrained power grid optimization results

Table 3.4 shows the lifetime constrained power grid optimization with our self-generated power grid networks. We remark that the IBM power grid circuits may not have the desired immortal wires suitable for the demonstration and it is more convenient to use self-generated power supply grids because we need to test our optimizer under different testing conditions. These network topologies are simple, the name pg10×10 means that the circuit consists of 10 rows and 10 columns power grid strips, in other words, there are 20 interconnect trees in total. So the size of the circuit in terms of nodes is approximately equal to # of rows × # of columns. In this table, column 3 shows the number of nucleated trees that stall the optimization. Note that nucleated wires can be the failed wires or wires that have the potential to fail (its final lifetime may be longer than the target lifetime even though its $V_E - V_{cathode}$ is larger than the critical voltage). If we have to relax the EM constraint, the maximum V_{crit} will be presented in column 6 ($V_{crit-max}$ *(V)*), meanwhile, the shortest lifetime after optimization is presented in column 5 ($t_{life-min}$ *(yr)*).

Table 3.4: EM lifetime constrained power grid networks for self-generated power grid networks.

circuit	# tree	# nuc-wires	before	after		CPU time	area reduced
			$t_{life-min}$ (yr)	$t_{life-min}$ (yr)	$V_{crit-max}$ (V)	(s)	(%)
pg5×10	15	1	-	immortal	-	1.84	76.51
pg10×10	20	5	80.68	77.39	4.307×10^{-3}	11.95	38.29
pg30×50	80	11	5.53	19.88	5.61×10^{-2}	96.39	26.68
pg20×100	120	8	> 100	> 100	7.22×10^{-2}	117.52	46.62

From Table 3.4, we can see if there are no violations for the initial power grid networks, the optimization can be easily carried on. Sometimes, even with some violations, after a few P-V and P-I iterations, the violations can be eliminated and we can still obtain EM immortal solution. Note that the area improvement strongly depends on the original

layouts, thus the absolute values of the reduced area are not that important. We notice that there are several nucleated wires in each case, which means that we do not check their EM during the optimization process as we treat them as failed wires. For instance, for *pg5×10* case, there exists a nucleated wire before optimization. We won't check the lifetime until the optimization finishes, however, after optimization, no lifetime calculation is needed as the network becomes immortal.

As for the *pg30×50* case, before optimization, the shortest lifetime $t_{life-min}$ is 5.53 years, which violates the lifetime constraint. After the optimization, the lifetime was improved to 19.88 years. As we can see from this example, the lifetime can be extended while the area can still be saved. The reason is that we let some wires fail or relax, but the failure of those wires (resistance change or open circuits) will be compensated by properly sizing of other wires to meet the lifetime requirement due to the redundant structure design of power grid networks. It may lead to a wider width, but the overall area of all the wires can be reduced. This further demonstrates the superior advantage of the lifetime constrained method over the immortality constrained method.

For other cases, even though they have a few nucleated wires, their lifetime is very long. After optimization, this is still the case. For most cases, if $V_E - V_i \gg V_{crit}$ in the initial condition, it is difficult to optimize successfully for the first time. With our lifetime constrained optimization flow, the optimization results can always be achieved after several iterations so that the lifetime target can be met.

3.6 Summary

In this chapter, we first presented a new power grid network sizing technique based on the voltage-based EM immortality check method for general multi-segment interconnect wires and the physics-based EM assessment technique for fast time to failure analysis. We showed that the new power grid optimization problem subject to the voltage IR drop and new EM constraints can still be formulated as an efficient sequence of linear programming problem. The new optimization method will ensure that none of the wires fail if all the constraints are satisfied. In addition, we improved the optimization method with reservoir branch insertion, which helps make the initial power grid network more robust. To mitigate the overly conservative nature of the optimization formulation, we further considered EM-induced aging effects on power grid networks for a target lifetime and then presented an EM lifetime constrained optimization method which allows some short-lived wires to fail and optimizes the rest of the wires. Numerical results on a number of IBM and self-generated power grid networks showed that the new methods can effectively reduce the area of the power grid networks while ensuring reliability in terms of immortality or target lifetime, which is not the case for the existing current density constrained power grid optimization methods.

Chapter 4

Robust power grid network design considering EM aging effects

4.1 Saturation volume-based EM immortality check

EM is a material migration process caused by an electrical field. Due to momentum exchange between atoms, hydrostatic stress is generated inside the embedded metal wire during the migration process. Before the hydrostatic stress reaches the critical level, the atomic flux flowing caused by electron flow from the cathode to the anode can still balance with the atomic flux caused by the inhomogeneous distribution of hydrostatic stress. When the stress reaches the critical level, a void and a hillock caused by atom migration will be formed at the cathode end and anode end, respectively. However, if the hydrostatic stress cannot reach the critical level, the void will not be formed or nucleated. The case where a void cannot be formed is called EM nucleation phase immortal.

After a void is formed, it will keep growing until saturation. Void saturation happens when two kinds of flux balance with each other. One is the flux of atoms previously located in metal which is consumed by the growing void and the other is the back flux of atoms generated by a gradient of growing stress. If the void volume is smaller than the volume of the intersection which is recognized as critical volume $\mathcal{V}_{crit}$, current can still flow through the copper wire since the wire cross-section is not blocked by the void. Once the void size is large enough and occupies the wire cross-section, current has to go through the liner whose resistivity is much higher than copper and the resistance of the wire will increase, indicating that the wire enters the growth phase. As can be seen, if the saturation volume is smaller than the critical volume, the wire is still immortal even if a void is formed. This case is called EM incubation phase immortal.

A model predicting the saturation volume $\mathcal{V}_{sat}$ was proposed in [51]:

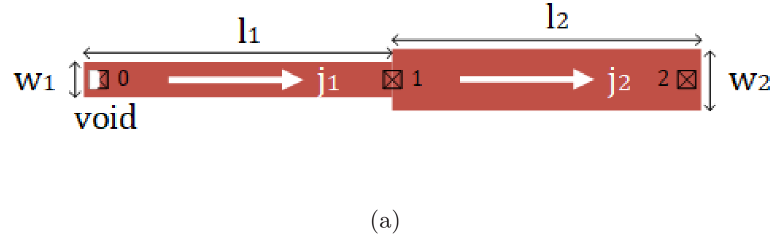
$$\begin{aligned} \mathcal{V}_{sat} &= \sum_i \mathcal{V}_{sat,i} = h \times \sum_i \mathcal{A}_{sat,i} \\ &= h \times \sum_i [(-2\sigma_{c,i} + \frac{j_i l_i \rho e Z}{\Omega}) \times \frac{l_i w_i}{2B}] \\ &= h \times \sum_i [(-2\sigma_{c,i} + \frac{V_i e Z}{\Omega}) \times \frac{l_i w_i}{2B}] \end{aligned} \tag{4.1}$$

where h is the thickness of the wire and $\mathcal{V}_{sat,i}$, $\mathcal{A}_{sat,i}$, $\sigma_{c,i}$, j_i, l_i and w_i represent the contribution to void volume, contribution to void area, stress at the cathode, current density, length, and width of the ith segment, respectively. For the segment in which a void has been nucleated, $\sigma_{c,i}$ is 0 on the cathode. Except for the segment with the void, the steady-state stress on the cathode of the other segments is the same as the anode of the segment connected to them.

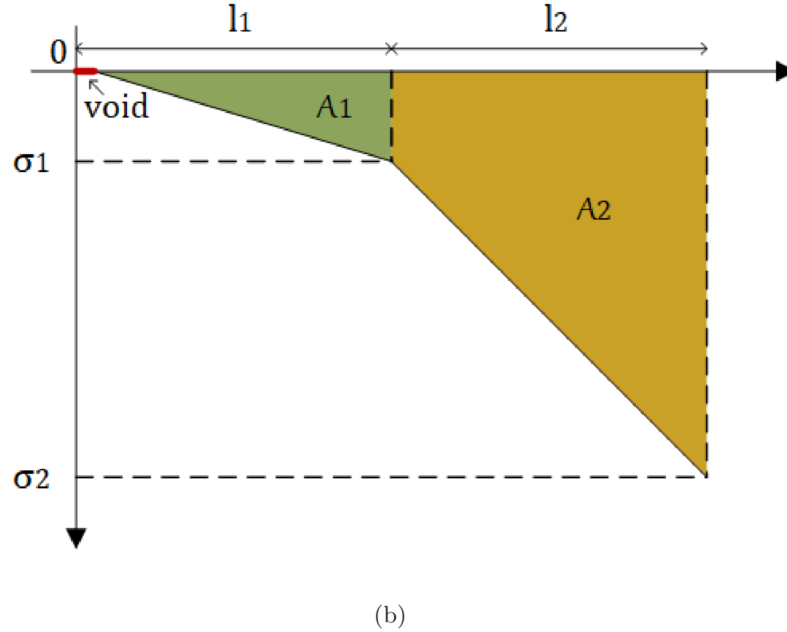
As shown in Eq. (4.1), voltage and width on each branch i contribute to the void volume, so the saturation void volume can be adjusted by modifying the voltage and width of the branches in order to make the wire incubation phase immortal.

Fig. 4.1 uses a two-segment wire to illustrate this method. In Fig. 4.1(a), the direction indicates electron flow. Here, stress at node 1 and node 2 can be calculated as

$$\begin{aligned}\sigma_1 &= 0 - \frac{(\mathcal{V}_1 - 0)eZ}{\Omega} = -\frac{j_1 l_1 \rho e Z}{\Omega} \\ \sigma_2 &= -\sigma_1 - \frac{(\mathcal{V}_2 - \mathcal{V}_1)eZ}{\Omega} = -\frac{(j_1 l_1 + j_2 l_2)\rho e Z}{\Omega}\end{aligned} \tag{4.2}$$



(a)



(b)

Figure 4.1: (a) A two-segment wire; (b) stress integration area of a two-segment wire.

Fig. 4.1(b) shows the stress at steady state versus branch length. $\mathcal{A}_1$ and $\mathcal{A}_2$ in the figure are the contributions of the two branches on void area, which indicate that the stress on both branches has a contribution to the void formation.

In addition, critical volume $\mathcal{V}_{crit}$ is defined as

$$\mathcal{V}_{crit} = h \times w \times d \tag{4.3}$$

where h is the thickness of the wire, w is the wire width and d is the via diameter. For wide wires, multiple vias may be applied, then d is equivalent to w. In this case, we can assume that $\mathcal{V}_{crit}$ is proportional to w^2.

In order to determine whether the wire is EM incubation phase immortal or not, the saturation volume is compared with the critical volume, more specifically, the wire is incubation phase immortal if

$$\mathcal{V}_{sat} < \mathcal{V}_{crit} \tag{4.4}$$

4.2 New EM immortality constrained optimization for multi-segment interconnects

In this section, we present a new EM immortality constrained power grid wire sizing method for multi-segment interconnect wires. Compared to previous work [72, 73], the new EM constraint is less conservative and can optimize more hard-to-optimize problems while ensuring all the wires are EM immortal.

4.2.1 New EM immortality criteria

EM conditions for multi-segment wires

Fig. 4.2 shows the void formation in a power grid network. We consider the following three cases.

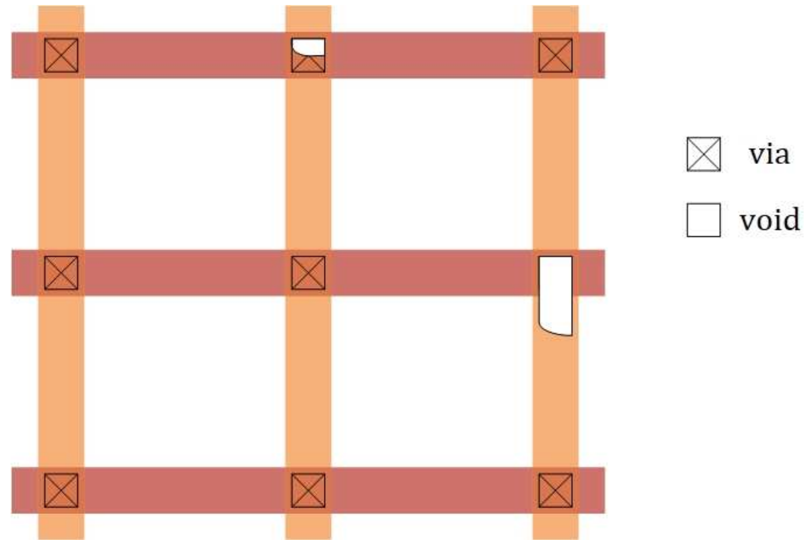


Figure 4.2: Void formation examples in a power grid network.

Case 1 In the left vertical wire, no void is formed, which means the stress at the cathode does not exceed the critical stress in the nucleation phase, thus the wire is EM immortal. We call it *EM nucleation phase immortal*.

Case 2 In the middle vertical wire, a void is nucleated, but it does not fully cover the via, EM failure process ends at the incubation phase in which no meaningful wire resistance is observed. Note that the via resistance may change due to void formation, but we do not take via resistance into consideration here. In conclusion, the void is saturated before

reaching the critical volume and the wire is still considered EM immortal. We call it *EM incubation phase immortal*.

Case 3 In the right vertical wire, a void is nucleated at the cathode, after the incubation phase, it fully covers the via, initiating the true growth phase. In this phase, the resistance starts increasing as current starts to flow through the more resistive barriers of the copper wire. When it increases to the critical level, such as 10% or other defined thresholds, the wire can be considered to be failed. We deem this kind of wire *EM mortal*.

New EM immortality check

In Chapter 3.1, EM immortality check criteria have been presented for general multi-segment interconnects. This voltage-based EM or *VBEM* method, in contrast to the traditional current density-based method, formulated the new criteria in terms of nodal voltages. Specifically, this EM immortality condition of a multi-segment interconnect is described as

$$V_E < V_{crit,EM} \tag{4.5}$$

However, this condition only checks void nucleation (Case 1). It fails to consider the case where a void is nucleated in a wire but the saturation void volume is less than the critical volume (Case 2).

Given a multi-segment wire m, the new EM immortality criteria first checks *VBEM*. If it passes, then the wire is immortal. Otherwise, we compute the saturation void volume $\mathcal{V}_{sat}$ from Eq. (4.1) and perform a saturation volume check using Eq. (4.4). If Eq. (4.4) satisfies, the wire is still considered to be immortal. Otherwise, it is mortal,

a more complicated analysis of transient hydrostatic stress evolution is needed to evaluate the time to failure.

Note that EM nucleation phase immortal indicates that the stress will not reach the critical level. No void is formed does not mean that the saturation volume is zero, however, saturation volume makes sense only after a void is nucleated.

We remark that the EM models applied in our work are for general cases, we will consider process variation with the Monte Carlo method in the future.

4.2.2 Problem formulation

Let $G = \{N, B\}$ be a power grid network with n nodes $N = \{1, ..., n\}$ and b branches $B = \{1, ..., b\}$. Each branch i in B connects two nodes i_1 and i_2 with current flowing from i_1 to i_2. l_i and w_i are the length and width of branch i, respectively. ρ is the sheet resistivity. The resistance r_i of branch i is

$$r_i = \frac{V_{i_1} - V_{i_2}}{I_i} = \rho \frac{l_i}{w_i} \tag{4.6}$$

Objective function

The total routing area of a power grid network is expressed in terms of nodal voltages, branch currents, and branch lengths as follows:

$$f(V, I) = \sum_{i \in B} l_i w_i = \sum_{i \in B} \frac{\rho I_i l_i^2}{V_{i1} - V_{i2}} \tag{4.7}$$

Constraints

The constraints that need to be satisfied for a reliable power grid network are shown as follows.

1. **Tree-related constraints**

   Equal width constraints For typical chip layout designs, branches within an interconnect tree should have the same width, i.e., $w_i = w_k$.

   $$\frac{V_{i1} - V_{i2}}{l_i I_i} = \frac{V_{k1} - V_{k2}}{l_k I_k} \tag{4.8}$$

   Note that typically an interconnect tree is just a power grid stripe and equal width is a common assumption for general power grid networks.

   New EM immortality constraints for multi-segment interconnects As described before, for a multi-segment interconnect m, we consider two EM immortality constraints: *the EM nucleation immortal constraint* (4.5) and *EM incubation phase immortal constraint* (4.4). We will select one of them based on the stress conditions. We remark that in constraints (4.5) and (4.4), both the *EM voltage* $V_{E,m}$ and the saturation volume $\mathcal{V}_{sat,m}$ are linear functions of the nodal voltages of the interconnect wires, supposing that each branch has the same width. As a result, the constraints are still linear in terms of nodal voltages, which is a requirement for the sequence of linear programming method.

2. **Other constraints**

   Voltage IR drop constraints A threshold voltage is required to guarantee proper logic operation

   $$V_j > V_{min} \text{ for power network} \tag{4.9}$$

Minimum width constraints Usually different layers have different requirements for the width of the wire segments

$$w_i = \rho \frac{l_i I_i}{V_{i1} - V_{i2}} \geq w_{i,min} \tag{4.10}$$

KCL For each node j, we have

$$\sum_{k \in B(j)} I_k = 0 \tag{4.11}$$

where $B(j)$ is the set of branches incident on node j.

4.2.3 Relaxed two-step sequence of linear programming solution

The power grid optimization aims to minimize objective function (4.7) subject to constraints (4.5), (4.4), (4.9)-(4.11). The problem is a constrained nonlinear optimization problem.

Note that $V_{E,m}$ is a function of both nodal voltage and area of a wire. According to objective function (4.7), the area of a wire segment is a function of nodal voltages and branch current, therefore, $V_{E,m}$ is a nonlinear function. Take a three-segment wire similar to Fig. 4.1(a) as an example,

$$\begin{aligned} V_{E,m} =& \frac{a_1 \left(V_0 - V_{cat,m}\right) + \left(a_1 + a_2\right) \left(V_1 - V_{cat,m}\right)}{2 \left(a_1 + a_2 + a_3\right)} \\ &+ \frac{\left(a_2 + a_3\right) \left(V_2 - V_{cat,m}\right) + a_3 \left(V_3 - V_{cat,m}\right)}{2 \left(a_1 + a_2 + a_3\right)} \\ =& \frac{l_1 w_1 V_0 + \left(l_1 w_1 + l_2 w_2\right) V_1 + \left(l_2 w_2 + l_3 w_3\right) V_2 + l_3 w_3 V_3}{2 \left(l_1 w_1 + l_2 w_2 + l_3 w_3\right)} \\ &- V_{cat,m} \end{aligned} \tag{4.12}$$

where $V_{cat,m}$ is the cathode node voltage of the mth wire. If we have the equal width constraint (4.8), which indicates $w_1 = w_2 = w_3$, then the EM nucleation phase immortal

constraint (4.5) actually becomes a linear function of nodal voltage again.

$$V_{E,m} = \frac{l_1 V_0 + (l_1 + l_2) V_1 + (l_2 + l_3) V_2 + l_3 V_3}{2 (l_1 + l_2 + l_3)} - V_{cat,m} \tag{4.13}$$

For $\mathcal{V}_{sat,m}$, it is also nonlinear in terms of nodal voltages, assuming V_0 is the cathode node voltage,

$$\begin{aligned}\mathcal{V}_{sat,m} =& h \times \Bigg[\left(-2\sigma_0 + \frac{(V_1 - V_0) eZ}{\Omega}\right)\frac{l_1 w_1}{2B} + \Bigg(- 2\sigma_1 \\ & + \frac{(V_2 - V_1) eZ}{\Omega}\Bigg)\frac{l_2 w_2}{2B} + \left(-2\sigma_2 + \frac{(V_3 - V_2) eZ}{\Omega}\right)\frac{l_3 w_3}{2B}\Bigg] \\ =& h \times \frac{eZ}{2B\Omega}\left[(V_1 - V_0) l_1 w_1 + (V_1 + V_2 - 2V_0) l_2 w_2 \right. \\ & \left. + (V_2 + V_3 - 2V_0) l_3 w_3\right]\end{aligned} \tag{4.14}$$

However, considering $\mathcal{V}_{crit,m}$ is proportional to w^2, EM incubation phase immortality condition can be simplified as

$$\frac{eZ}{2B\Omega}\left[(V_1 - V_0) l_1 + (V_1 + V_2 - 2V_0) l_2 + (V_2 + V_3 - 2V_0) l_3\right] < w \tag{4.15}$$

With all these linear constraints, we can follow the relaxed two-phase iterative optimization process and apply the sequence of linear programming technique [72] to solve the relaxed problem in each phase.

Note that the equal width constraint aims for the optimization flow only, the segments of an interconnect can have different widths due to reservoir insertion [73]. Besides, sensitivity-based IR drop fix methods, which will be discussed in Chapter 4.4, can also lead to distinct segment widths.

Specifically, we have a voltage solving phase when all branch currents are assumed to be fixed and a current solving phase when all nodal voltages are fixed. Because the objective function in the voltage solving phase is nonlinear, we take the first-order Taylor's

expansion around the initial solution V^0 to get the linearized objective function.

$$g(V) = \sum_{i \in B} \frac{2\rho I_i l_i^2}{V_{i1}^0 - V_{i2}^0} - \sum_{i \in B} \frac{\rho I_i l_i^2}{\left(V_{i1}^0 - V_{i2}^0\right)^2} (V_{i1} - V_{i2}) \tag{4.16}$$

Since I_i is a constant, an additional constraint is added

$$\xi sign(I_i) \left(V_{i1}^0 - V_{i2}^0\right) \leq sign(I_i) (V_{i1} - V_{i2}) \tag{4.17}$$

where $\xi \in (0, 1)$ is a restriction factor, which will be selected by some trials and experiences, and $sign(x)$ is the sign function.

4.3 Comprehensive power grid optimization strategies

4.3.1 New EM immortality constrained power grid optimization

In the aforementioned optimization problem, the new EM immortality constrained power grid optimization starts with an initial feasible solution obtained from Chapter 2.2. Then we iteratively go through the voltage solving phase and current solving phase. The entire EM immortality constrained power grid network optimization procedure is summarized as Algorithm 4.

Algorithm 4 New EM immortality constrained power grid wire sizing algorithm

Input: Spice netlist G_I containing a power grid network.

Output: Optimized power grid network parameters.

1: /*Problem Setup*/

2: $k := 0$.

3: Compute the initial V^k, I^k from G_I.

4: **repeat**

5: /*P-V Phase*/

6: Perform new saturation volume-based EM immortality check flow, construct equal width constraints (4.8) and EM nucleation phase immortal constraints (4.5) or EM incubation phase immortal constraints (4.4).

7: Construct minimum width constraints (4.10) and linear companion constraints (4.17) with I^k.

8: $m := 1$.

9: Compute $V_m^k := \arg\min g\left(V^k\right)$ subject to constraints (4.9), (4.10), (4.5), (4.4), (4.8), and (4.17) by a sequence of linear programmings.

10: **while** $f\left(V_m^k\right) > f\left(V_{m-1}^k\right)$ **do**

11: Determine the search direction $d_m := V_{m-1}^k - V_m^k$.

12: Choose step size α for line search.

13: $V_{m+1}^k := V_m^k + \alpha d_m$.

14: $m := m + 1$.

15: **end while**

16: $V^{k+1} := V_m^k$.

17: /*P-I Phase*/

18: Construct equal width constraints (4.8) and minimum width constraints (4.10) with V^{k+1}.

19: Compute $I^{k+1} := \arg\min f\left(I^k\right)$ subject to constraints (4.10), (4.8) and (4.11).

20: $k := k + 1$.

21: **until** $\left|f\left(V^k, I^k\right) - f\left(V^{k-1}, I^{k-1}\right)\right| < \varepsilon$

At the beginning of each voltage solving phase (line 6), we check the EM immortality for all the interconnect trees. If there exists a wire, which is neither EM nucleation phase immortal nor EM incubation phase immortal, we consider the power grid cannot be optimized. Otherwise, one EM constraint is built for each wire for later optimization. Note that the length and width of a certain branch used in each iteration may differ, thus the constant part of the EM constraints may change. The new saturation volume-based EM immortality check flow is shown in Fig. 4.3.

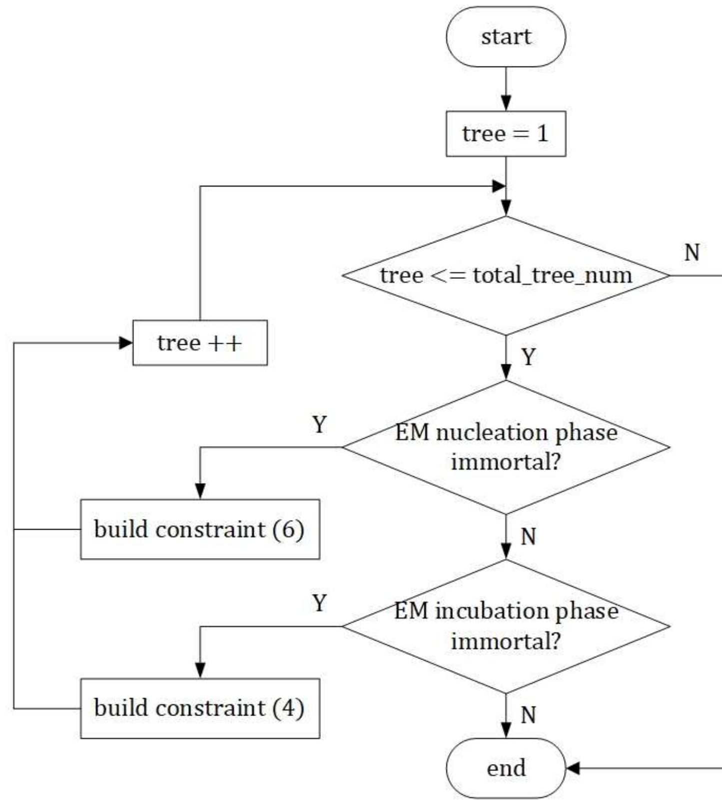


Figure 4.3: Saturation volume-based EM immortality check for power grid networks.

4.3.2 EM immortality constrained power grid optimization with wire pre-sizing

EM immortality requirement for all the interconnects at the initial stage sometimes can be too strict. This happens when the initially designed power grids do not meet the EM requirement. As discussed earlier, we check the EM immortality conditions at the beginning of the voltage solving phase and consider the power grid network cannot be optimized if neither immortality condition is satisfied, which will greatly reduce the optimization space.

To mitigate this problem, we present a wire pre-sizing strategy, which can be seen as a preprocessing stage before the power grid optimization. In other words, in order to enable the optimization, we keep the original topology but adjust the resistance/width of power grid stripes.

When a wire fails the immortality criteria, one idea is to size the width up to make the wire immortal subject to the design rules. When we increase the width of the wire segments or the interconnect tree (for all of its segments), we will increase the critical void volume, which is increased quadratically with width w as defined in Eq. (4.3). As a result, the wire may become incubation phase immoral based on Eq. (4.4). Besides, width increase will also reduce the current density and branch voltages, therefore, it may make the wire nucleation phase immortal based on Eq. (4.5). In our approach, we size up the failed wires so that one of the immortal conditions is met. The wire is sized by recomputing the branch voltages assuming that it will not affect other connected interconnect trees as a first-order approximation until one of the immortality conditions is met.

If the number of the wire segments need to be sized up is large, it indicates that the power grid network is not well designed in terms of EM in the first place and some structure change (e.g., adding more power strips) is probably needed first. Sizing up wire segments is trivial as long as design rules are allowed, which is an efficient way of preprocessing.

4.4 IR drop constrained power grid optimization

For power grid network design, the ultimate design constraint is still the IR drop at the target lifetime (e.g., 10 years), which means that we can allow some EM failures in the power delivery networks as long as IR drop constraints are met. In this section, we present two strategies to optimize the power delivery networks to meet IR drop constraints with relaxed EM constraints.

4.4.1 EM lifetime constrained power grid optimization

Previous methods use EM immortal constraints to ensure that none of the interconnect trees fail. However, such constraints may be conservative because some wires can have EM failures as long as the power grid networks still work in the target lifetime and it fails to consider the aging effect. Therefore, an EM aging-aware power grid wire sizing optimization method is applied in which some segments of interconnects will be allowed to fail or to increase resistance.

After performing an EM immortality check, if neither of the EM constraints is met for a certain interconnect tree, then the lifetime of the wire will be computed based on the fast EM lifetime estimation method [62]. If the calculated lifetime is smaller than the

target lifetime, the interconnect will be marked as failed and its EM constraint will not be considered anymore. Furthermore, if the void is formed in via-above interconnect, the via will be treated as disconnected, otherwise (via-below case), the increase in resistance of the segment will be computed [73].

On the other hand, if the calculated lifetime meets or exceeds the target lifetime, the saturation volume obtained from the previous optimization $\mathcal{V}_{sat,m,prev}$ for mth interconnect will become the new critical void saturation volume for the next optimization iteration, i.e., Eq. (4.4) becomes

$$\mathcal{V}_{sat,m,prev} > \mathcal{V}_{sat,m} \tag{4.18}$$

The rationale behind this is that we relax the incubation phase immortality constraint as it still leads to a satisfactory lifetime for this wire in the previous optimization. By adding this constraint, we expect that its lifetime will not change too much and still achieve the given lifetime goal after the follow-up optimizations. After either *resistance change*, or *wire disconnection*, or *constraint relaxation*, a new round of SLP optimization is carried out until the optimized chip meets the EM lifetime target. Compared with [73], there will be more sizing space with the saturation volume-based constraint and it will provide better optimization results.

4.4.2 Sensitivity-based localized power grids fixing

In this section, we present a large change sensitivity-based IR drop fixing method, which aims at enhancing EM reliability at late power grid network design stage, i.e., sign-off or electrical engineering order stages. As we know, both early failure and late failure

contribute to the voltage change. When an early failure occurs, the interconnect is disconnected from another one, which can be seen as an open circuit. In contrast, the resistance of late failure trees increases gradually and finally stops when the voids reach the saturation volume. As a result, the wire resistance can experience large changes (from zero to infinity) during the target lifetime. Note that the definition of power grids lifetime is different from wire lifetime. It refers to the time at which an EM-induced voltage failure is expected to happen.

Algorithm 5 presents the large change sensitivity-based fixing algorithm.

If there exist mortal wires after performing EM immortality check, we deem the power grid network is a mortal one. For mortal power grids at design time, we first perform the IR drop simulation. The MNA equation for the power grid network at t_0 is represented as

$$\mathbf{G}(t_0)\mathbf{v}(t_0) = \mathbf{i} \tag{4.19}$$

where $\mathbf{G}$ is the admittance matrix of the network, $\mathbf{v} = [V_1, ..., V_n]^T$ is the vector of resulted node voltages and $\mathbf{i}$ is the attached current sources. The conductance of branch i is

$$g_i = \frac{w_i}{\rho l_i} \tag{4.20}$$

Therefore, the elements of $\mathbf{G}$ can be expressed as

$$G_{ij} = \begin{cases} -\sum_{k=1}^{N} g_k, & i = j \text{ and } (k_1 = i \text{ or } k_2 = i) \\ g_k, & i \neq j \text{ and } (k_1 = j \text{ or } k_2 = j) \\ 0, & i \neq j \text{ and } k_1 \neq j \text{ and } k_2 \neq j \end{cases} \tag{4.21}$$

Second, the coupled EM-IR drop analysis for the power grid network is performed through *EMspice* [53, 2]. The new MNA equation at the target lifetime t_{target} can be

Algorithm 5 Large change sensitivity-based power grid fixing algorithm

Input: Spice netlist G_I containing a vulnerable power grid network.

Output: Robust power grid network spice netlist G_O.

1: Perform IR drop simulation and obtain $\mathbf{G}^k(t_0)$ and $\mathbf{v}^k(t_0)$ for the power grid network at initial time t_0 .

2: Perform EM-IR co-simulation and obtain $\mathbf{G}^k(t_{\text{target}})$ and $\mathbf{v}^k(t_{\text{target}})$ at time point t_{target}.

3: **while** $\mathbf{v}^k(t_{\text{target}}) \leq V_{\text{threshold}}$ **do**

4: Calculate the difference of m resistor widths $\Delta\mathbf{w}^k$ using Eq. (4.23) and determine the sensitivity matrix $\mathbf{S}^k_{n\times m}$.

5: Obtain the error vector $\boldsymbol{\epsilon}^k$ based on the threshold voltage $V_{\text{threshold}}$ and $\mathbf{v}^k(t_{\text{target}})$.

6: Use gradient descent method Eq. (4.33) to update the width $\mathbf{w}^k(t_0)$ and widen the branch i with $w_i^{k+1}(t_0)$ which satisfies $w_i^{k+1}(t_0) > w_i^k(t_0)$.

7: **if** the updated branch width does not meet the design rules **then**

8: Set the branch width to the maximum allowable value and mark it as unadjustable.

9: **end if**

10: Perform IR drop simulation and obtain $\mathbf{G}^k(t_0)$ and $\mathbf{v}^k(t_0)$ for the power grid network at initial time t_0 .

11: Perform EM-IR co-simulation and obtain $\mathbf{G}^k(t_{\text{target}})$ and $\mathbf{v}^k(t_{\text{target}})$ at time point t_{target}.

12: **end while**

expressed as

$$\mathbf{G}(t_{\text{target}})\mathbf{v}(t_{\text{target}}) = \mathbf{i} \tag{4.22}$$

Assume that m resistor $(g_1, g_2, \ldots, g_m)$ values would change due to void formed at the cathode. Then the difference of the corresponding effective width $\Delta\mathbf{w} = [\Delta w_1, \Delta w_2, \ldots, \Delta w_m]^T$ between t_0 and t_{target} time points are given by

$$\Delta w_i = \Delta g_i \rho l_i = [g_i(t_{\text{target}}) - g_i(t_0)]\rho l_i \tag{4.23}$$

Finally, we only change the i-th width with increment Δw_i at one time to calculate the voltage increments $\Delta\mathbf{v} = [\Delta v_1, \Delta v_2, \ldots, \Delta v_n]$ compared to the original network. When changing the i-th resistor, the $\mathbf{G}(t_0)$ matrix is modified by adding the perturbation matrix $\Delta g_i \mathbf{p}_i \mathbf{q}_i^T$, which is expressed as

$$\mathbf{G}\mathbf{v} = \left(\mathbf{G}(t_0) + \Delta g_i \mathbf{p}_i \mathbf{q}_i^T\right)(\mathbf{v}(t_0) + \Delta\mathbf{v}) = \mathbf{i} \tag{4.24}$$

where $\mathbf{p}_i$ and $\mathbf{q}_i$ are $n \times 1$ vectors. Their elements are given by

$$[\mathbf{p}_i]_k = \begin{cases} 1, k = i_1 \\ -1, k = i_2 \\ 0, \text{otherwise} \end{cases} \quad [\mathbf{q}_i]_k = \begin{cases} -1, k = i_1 \\ 1, k = i_2 \\ 0, \text{otherwise} \end{cases} \tag{4.25}$$

Based on Eq. (4.24), the increments of voltage can be determined by

$$\Delta\mathbf{v} = -(\mathbf{G}(t_0) + \Delta g_i \mathbf{p}_i \mathbf{q}_i^T)^{-1} \Delta g_i \mathbf{p}_i \mathbf{q}_i^T \mathbf{v}(t_0) \tag{4.26}$$

This equation involves the inverse of an $n \times n$ matrix and can be rewritten as [23]

$$\Delta\mathbf{v} = -\mathbf{G}(t_0)^{-1}\mathbf{p}_i(\Delta g_i^{-1} + \mathbf{q}_i^T \mathbf{G}(t_0)^{-1}\mathbf{p}_i)^{-1}\mathbf{q}_i^T \mathbf{v}(t_0) \tag{4.27}$$

Therefore, we only need to calculate the inverse of $\mathbf{G}(t_0)$ once to obtain the sensitivity $[\frac{\Delta v_1}{\Delta w_i}, \frac{\Delta v_2}{\Delta w_i}, \ldots, \frac{\Delta v_n}{\Delta w_i}]^T$. By repeating the above procedure m times, we can compute the sensitivity matrix $\mathbf{S}_{n\times m}$ as

$$\mathbf{S}_{n\times m} = \begin{bmatrix} \frac{\Delta v_1}{\Delta w_1} & \frac{\Delta v_1}{\Delta w_2} & \cdots & \frac{\Delta v_1}{\Delta w_m} \\ \frac{\Delta v_2}{\Delta w_1} & \frac{\Delta v_2}{\Delta w_2} & \cdots & \frac{\Delta v_2}{\Delta v_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\Delta v_n}{\Delta w_1} & \frac{\Delta v_n}{\Delta w_2} & \cdots & \frac{\Delta v_n}{\Delta w_m} \end{bmatrix} \tag{4.28}$$

Therefore, the relationship between $\Delta\mathbf{v}$ and $\Delta\mathbf{w}$ can be described by

$$\Delta\mathbf{v} = \mathbf{S} \times \Delta\mathbf{w} \tag{4.29}$$

To solve $\Delta\mathbf{w}$, we have

$$\begin{aligned} \mathbf{S}^T\Delta\mathbf{v} &= \mathbf{S}^T\mathbf{S} \times \Delta\mathbf{w} \\ \Delta\mathbf{w} &= (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\Delta\mathbf{v} \end{aligned} \tag{4.30}$$

Similar to constraint (4.9), to meet the operating requirements of the VLSI circuit, the voltages at all nodes should be no less than the threshold voltage $V_{\text{threshold}}$ within a certain lifetime. Therefore, our destination is to minimize the error vector

$$\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]^T \tag{4.31}$$

where

$$\epsilon_i = \begin{cases} 0, & V_i(t_{\text{target}}) \geq V_{\text{threshold}} \\ V_{\text{threshold}} - V_i(t_{\text{target}}), & V_i(t_{\text{target}}) < V_{\text{threshold}} \end{cases} \tag{4.32}$$

Then we can optimize the problem using the gradient descent method

$$\mathbf{w}^{k+1}(t_0) = \mathbf{w}^k(t_0) + \alpha(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\boldsymbol{\epsilon} \tag{4.33}$$

where α is the step size and can be selected by some trials. Note that the required width $\mathbf{w}$ to be changed maybe also subject to the design rules as there is an upper bound for the widths.

4.5 Experimental results and discussions

The presented EM-aware constrained power grid optimization is implemented in C/C++. IBM power grid benchmarks [43] are used to test our work, and we also have a few synthesized IBM-format power grid networks so that different kinds of EM immortality constraints can be tested and verified.

4.5.1 New EM immortality constrained power grid optimization results

Table 4.1 compares the results of the new optimization method considering saturation volume with the results of the existing power grid optimization method checking EM nucleation phase immortality only [72].

Table 4.1: EM immortality and lifetime constrained power grid optimization considering void saturation volume results.

circuit	# tree	# nuc-wire		area	w/o saturation volume [72]	w saturation volume
		(b)	(e)	(mm^2)	area reduced (%)	area reduced (%)
ibmpg2	462	0	0	60.38	77.55	77.55
PG1	128	0	0	40.21	34.34	34.34
PG2	38	0	9	0.50	28.78	38.33
PG3	9	1	3	0.031	can't finish (mortal wires (II))	53.26
PG4	12	1	1	0.030	can't finish (mortal wires)	25.21
PG5	20	10	10	0.017	can't finish (mortal wires)	14.84
PG6	20	2	2	0.23	27.39	28.41
PG7	80	11	11	1.28	26.68	29.76

In the optimization process, we assume that all the branches have the same width in one tree but different trees can have different widths. For a fair comparison, we only allow the difference in the EM constraints. In Table 4.1, column 1 to 5 list the power grid network benchmarks (*circuit*), the number of interconnect trees (*# tree*), the number of nucleated wires at the beginning (*# nuc-wires (b)*), the number of nucleated wires at the end (*# nuc-wires (e)*), and the original area (*area* (mm^2)). Column 6 and 7 report the reduced chip area ratio (*area reduced (%)*) with respect to the original area for the two methods with EM nucleation phase immortal only constraint (*w/o saturation volume*) and the new EM immortality criteria (*with saturation volume*) respectively. Note that for the two methods, all the wires are EM immortal after optimization.

As we can see, for the *ibmpg2* example, which has 462 immortal trees at the beginning, the original area is 60.38mm^2. After 2 iterations, the area can have a 77.55% reduction without any EM violation. For *PG2*, all the wires are EM nucleation phase immortal at first, after 2 iterations, although it has 9 nucleated wires, all of them are EM incubation phase immortal. Compared with the previous work [72] which has about 28.78% reduction, the new method has better performance in terms of area reduction, while still ensuring EM immortality. Previous work can not size *PG3* because it has nucleated wires in the beginning. Since the nucleated wires can still pass the saturation volume check (it is incubation immortal (II)), the power grid network can be sized properly with the new method. Note that the area improvement strongly depends on the original layouts, thus the absolute values of the reduced area are not that important.

We further observe that the new method typically leads to more area reduction compared to the previous method. The reason is that the new EM immortality criteria are less conservative and thus it effectively allows the *EM voltage* to exceed the *critical EM voltage* by using the *EM incubation phase immortal* constraint. In other words, a larger current density that can nucleate a void is allowed in the new method while ensuring wire resistance is not affected. As a result, we can conclude that the new method successfully enables the optimization of power grids, which would have been impossible in the previous method due to the conservative immortality constraint, in addition to leading to further area reduction in power grids compared to the previous method.

4.5.2 EM immortality constrained power grid optimization results with wire pre-sizing

The results of the EM immortality constrained power grid optimization with sizing up are also shown in Table 4.1. There exist some benchmarks which do not satisfy the EM requirement in the initial situation, in this way, we check if the interconnects can be sized up to become incubation phase immortal. In the *PG4* example, we find that there is one nucleated wire which fails both immortality constraints in the beginning. As discussed in Chapter 4.3, after width adjustment, it becomes incubation phase immortal. For *PG5*, half of the interconnects are mortal at first, and they all become nucleation phase immortal after wire pre-sizing. In this way, the optimization process can finish successfully.

4.5.3 EM lifetime constrained power grid optimization results

PG6 and *PG7* in Table 4.1 demonstrated the lifetime constrained optimization case. In the *PG6* example, the lifetime of the nucleated wires is longer than 100 years, thus the difference between the previous method and the new method is the relaxed constraint. In the *PG7* example, before optimization, the shortest lifetime among the 11 nucleated wires is 5.53 years, which violates the lifetime constraint. After the optimization, the lifetime was improved to 17.02 years. As we can see from this example, the lifetime can be extended while the area can still be saved. The reason is that the open circuit or resistance change of the short-lived wires will be compensated by properly sizing the other wires to meet the lifetime requirement due to the redundant structure design of power grid networks. It may lead to a wider width, but the overall area of all the wires can be reduced. This further demonstrates the superior advantage of the lifetime constrained method over the immortality constrained method. For the cases with very long lifetime wires, even though they may have a few nucleated wires, after optimization, this is still the case. Besides, compared with the previous method, the void saturation volume-based method can lead to more area reduction.

4.5.4 Sensitivity-based localized power grid fixing results

The sensitivity-based localized power grids fixing results are demonstrated in Table 4.2. The power grid designs come from the power grids of the Cortex-M0 DesignStart processor, which is a 32-bit processor that implements the ARMv6-M architecture. The

processor is synthesized using Synopsys Design Compiler and we simply modified some of the interconnects for our experiments.

Table 4.2: Sensitivity-based localized fixing results.

circuit	# tree	# violations at t_{target}	# mortal wires		# modified bch	area increase	total time
			(b)	(e)			
ARM-PG1	64	11	3	3	2	0.0438%	150.27s
ARM-PG2	64	32	5	5	3	0.0352%	171.65s
ARM-PG3	64	87	4	4	2	0.0614%	377.18s
ARM-PG4	128	21	7	6	4	0.0348%	522.83s

We assume that the widths of the original power grids have already been set to their minimums and thus cannot be reduced. Column 1 to 4 list the power grid network benchmarks (*circuit*), the number of interconnect trees (*# tree*), the number of IR drop violations at the target lifetime (*# violations at t_{target}*), and the number of mortal wires at the beginning (*# mortal wires (b)*). Column 5 to 8 report the number of mortal wires at the end (*# mortal wires (e)*), the number of modified branches (*# modified branches*), the increased chip area ratio (*area increased (%)*) with respect to the original area, and the total simulation time (*total time (s)*).

With the localized fixing method, we are able to meet the power grid lifetime by only widening a few branches. In the *ARM-PG1* example, the lifetime of the whole power grids is 8 years and the maximum voltage at t_{target} is $10.022\%V_{dd}$. There are 3 mortal wires and all of them have a lifetime of less than 10 years, in other words, there are 3 branch resistance changes at t_{target}. Note that the lifetime definitions of wire and power grid network are different. After performing the sensitivity analysis, we get the information that one of the wires needs to modify. The power grid meets the lifetime target with only one

iteration. The largest resistance change does not mean it contributes the most to voltage violations. For *ARM-PG3* who has 87 violations, the largest resistance change occurs at tree 33. However, according to Eq. (4.30) tree 1 and tree 31 should be modified, and tree 1 has the smallest resistance among the 4 mortal wires. In addition, *ARM-PG4* illustrates that the number of mortal wires can be reduced after the fixing process, where the mortal wires are less at the end than at the beginning. The total fixing time mainly depends on the iteration number and the simulation time from initial to 10 years to get the EM-induced IR drop.

The sensitivity-based localized power grid fixing method is better performed at the later design stages, we suppose that the power grid is well designed and meets the IR drop requirement at the start. Usually, the grid can be easily fixed to have a lifetime of 7 years or more. If the lifetime is too short, then there will be much more violations at 10 years, and maybe difficult to fix with widening only a small set of branches. In this case, the interconnect widening method will be applied first.

4.6 Summary

In this chapter, we presented several power grid network design and optimization techniques to consider the electromigration effects for multi-segment interconnect wires. First, we considered a new EM immortality constraint due to EM void saturation volume, which will reduce conservativeness in the EM-aware on-chip power grid design. We showed that the new EM immortality constraint can be integrated into the existing programming-based power grid optimization framework. Second, to further mitigate the excessive conser-

vativeness of the immortality constrained optimization methods, we explored three strategies: we first sized up failed wires to meet one of the immorality conditions subject to the design rules; next, we considered the EM-induced aging effects on the power grid networks for a target lifetime, which allows some short-lived wires to fail and optimizes the rest of the wires. Third, we presented a large change sensitivity-based optimization scheme to perform localized fixing based on the coupled EM-IR drop analysis method for power grid networks. Numerical results on a number of IBM-format power grid networks demonstrated that the new method can reduce more power grid area compared to the existing EM immortality constrained optimizations. Furthermore, the new method is able to optimize power grids with nucleated wires for the first time. Results also showed the sensitivity-based localized power girds fixing can fix EM-induced IR drop violations in a few minutes for synthesized power grid networks from ARM core designs.

Chapter 5

GridNet: Fast data-driven EM-induced IR drop prediction and localized fixing for on-chip power grid networks

5.1 Related works

5.1.1 Machine learning-based IR drop analysis and estimation

IR drop analysis is concerned with voltage drop estimation from given current or power sources, which can be time-varying for dynamic analysis. A number of numerical techniques have been well developed and can perform IR drop analysis well on power grids,

such as hierarchical methods, random walk methods, Krylov-subspace methods, multi-grid techniques, and vector-less verification methods.

To further speed up the IR drop analysis, several machine learning-based IR drop estimation/prediction methods have been proposed [39, 25, 30, 66]. Those methods typically aim to replace the standard full-chip IR drop analysis tool such as ANSYS RedHawk, via data-based learning and feature selections. For instance, Lin *et al.* [39] proposed full-chip dynamic IR drop analysis based on some power and physical features extracted from cells and layouts. Fang *et al.* [25] tried to improve the scalability by training the models for localized regions of the layout. Xie *et al.* [66] proposed a CNN-based model transferable across different designs that is able to incorporate design-dependent features during preprocessing. Ho *et al.* [30] focused on incremental IR drop prediction and mitigation. It uses more electrical and physical features for the training based on the gradient boosting framework.

5.1.2 EM-induced IR drop analysis and fixing works

Since EM failure may lead to wire resistance increase and even open circuits, it can cause an increase of IR drops over time. As a result, it is very important to perform the EM analysis and eventually EM-induced IR drop analysis towards the user-specified target lifetime. Many full-chip EM analyses for power grid networks have been proposed [33, 13, 49, 48]. These methods can predict the EM lifetime of the power grid and obtain failed trees. Specifically, Huang *et al.* proposed the first physics EM model-based full-chip EM analysis method [33, 34]. This method indeed considers the interaction between the EM and IR drops of power grids, but the compact EM model is less accurate. Chatterjee *et al.* proposed a

finite difference method (FDM) based full-chip EM analysis tool [13, 49] to get better accuracy. However, such method still primarily considers the EM stress without considering impacts from wire resistance changes of power grid networks. Cook *et al.* proposed a finite difference analysis method, which was accelerated by the Krylov subspace-based reduction technique [19]. This method can be applied to general multi-segment interconnect wires with time-varying current and temperature. However, this method still considers only the EM stress and ignores the EM and IR drop interaction in the power grid networks.

Recently, Sun *et al.* [53] proposed a full-chip EM-induced IR drop analysis, which considers the dynamic interplay between the hydrostatic stress and electronic current/voltage in a power grid network. This method solves the coupled time-varying partial differential equations in the time domain accurately and obtains the stress evolution in multi-segment interconnect trees. It is compatible with the synthesized power grid networks from commercial design tools and can show the resulted IR drop and EM failure hotspots at the target lifetime. However, the simulation can be very slow and hence not practical to use for fast EM fixing. This is one of the major issues that motivate our work.

At the same time, there are some works proposed on wire segment sizing of power grid networks in order to fix the EM failures or IR drops. Zhou *et al.* [74, 70] proposed a power grid network sizing method based on the multi-segment EM immortality check criteria. It automatically considers all the wire segments and their interactions in an interconnect tree. However, the EM immortality constrained optimization is still conservative as it requires all the interconnect trees to be immortal. Chang *et al.* [11] proposed a learning-based EM violation waiver system, which investigates every EM violation and takes an

expert decision to either ignore the violation (waive-off) or resolve it (must-fix) in the design. However, this system is not able to take the violation fix action. Moudallal *et al.* [42] directly optimized the EM-induced IR drops on the time-varying power grid networks due to the EM aging process. This method is based on gradient descent optimization and aims to size the individual wires to meet the target IR drop criteria, however, a large amount of computation is required.

5.2 Fast data-driven incremental EM-induced IR drop prediction

5.2.1 Overall workflow of the *GridNet* framework

Fig. 5.1 and Fig. 5.2 demonstrate the overall workflow of our *GridNet* framework. The workflow consists of two phases: training and inference. The training phase is shown in Fig. 5.1, the yellow block shows how the power grids are generated. Then in the red block, we use *EMspice* [53] to predict the EM-induced IR drop for synthesized power grid network using coupled EM-IR analysis. In the blue block, *GridNet* receives the EM-induced voltage from 0 to T_{target} aging years as well as the power grid. It extracts electrical and other information features, the training process is shown with dashed arrows. Fig. 5.2 illustrates the inference phase and the sensitivity-based power grid fixing flow. One of the two outputs from *GridNet* is the EM-induced voltage at all the nodes at a specific aging year. And the other is the sensitivity information: sensitivity of nodal voltages with respect to the input resistances. These resistances can be obtained as a by-product from the differentiable conditional generative adversarial network (CGAN) model as we will show

later. The sensitivity information will be utilized for fixing IR drop violations efficiently in the chip design flow. After incrementally updating the power grid, the new EM-induced voltage is predicted by the *GridNet* model. If IR drop violations remain unaddressed, the designer can just iteratively perform the same round of incremental prediction and fixing until all IR drop violations are fixed.

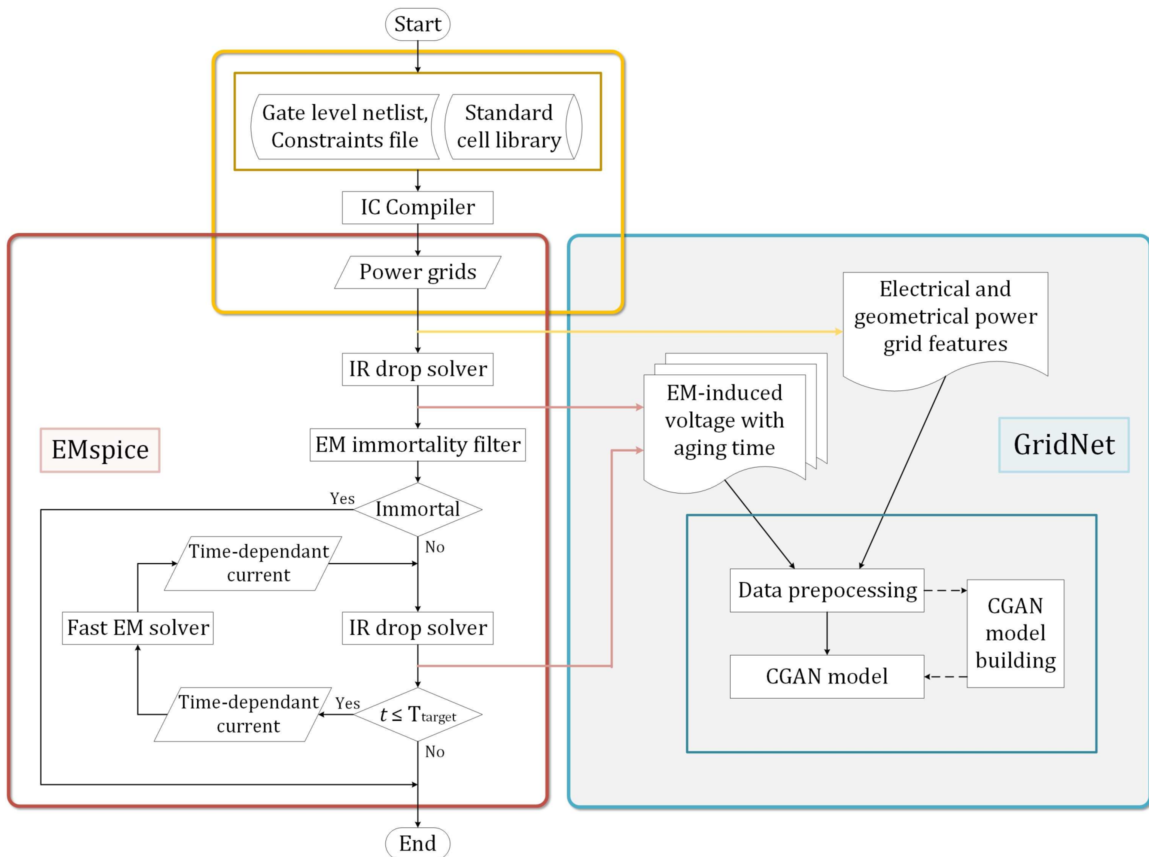


Figure 5.1: *GridNet* training flow.

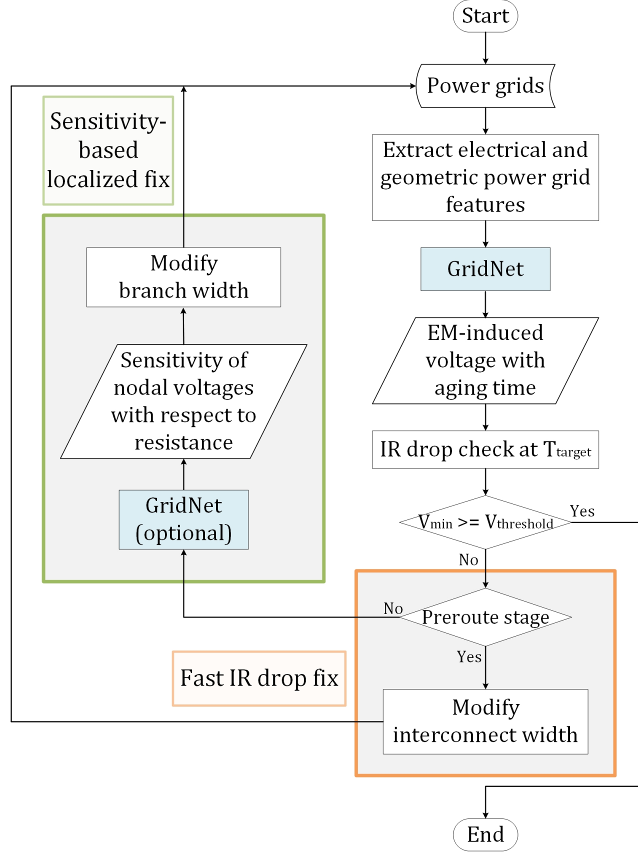


Figure 5.2: *GridNet* prediction flow and grids fix flow with *GridNet*.

5.2.2 Feature selections for *GridNet*

Given a mesh-structured power network, if we only look at the node voltage and input current sources, according to Eq. (2.9), we can formulate the time-varying model essentially as following

$$\mathbf{M}(t) \times u(t) = \mathbf{P}I(t), \tag{5.1}$$

where $\mathbf{M}(t)$ is a time-varying admittance matrix, $I(t)$ is a column vector whose elements are current and voltage sources, and $u(t)$ contains both nodal voltages and dependent current variables. As a result, for the DNN-based modeling, the input features should include both

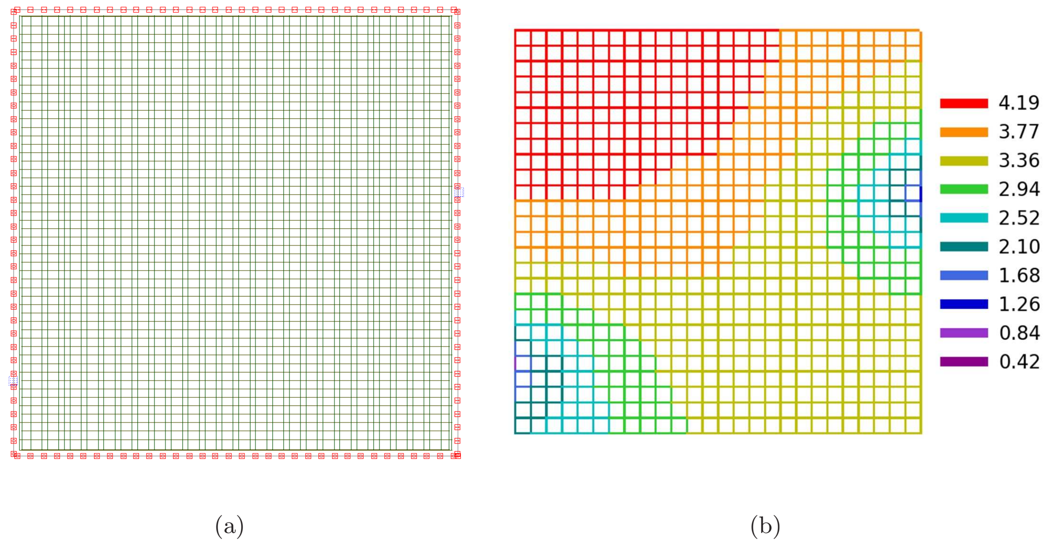
$I(t)$ and $\mathbf{M}(t)$, which can be represented by the resistance vectors of wire segments in the power grid networks. The resistance of a wire depends on its length and cross-sectional area that is proportional to wire width. Since we deal with mesh-structured power grids, the topology of wire connections is implicitly presented if all the wire resistance or features are pre-ordered (as a vector) based on some counting order. As a result, the *GridNet* model is able to deal with different workloads, i.e., $I(t)$ and initial wire resistances (different $\mathbf{M}$ at $t = 0$ under the same power grid structure).

5.2.3 Training data preprocessing and representation

The preprocessing extracts the electrical features and geometries from raw layouts. After the preprocessing, the workload samples will be represented in a customized scheme.

Data preprocessing

Given a specific design, Synopsys IC Compiler takes a synthesized gate-level netlist and a standard cell library as input, and then automatically creates the circuit layout. In the preroute (design planning) step, one important procedure is performing power network synthesis. As shown in Fig. 5.3(a), the power and ground network are generated based on the constraints that the user defines. It consists of V_{DD} power nets, V_{SS} ground nets, and two external supplies. The results are later used to examine the voltage drop, resistance and EM. Fig. 5.3(b) shows the voltage drop from the same power grid and the unit is mV. Since our goal is to obtain EM-induced IR drop which considers the aging effect, we dumped the



(a)

(b)

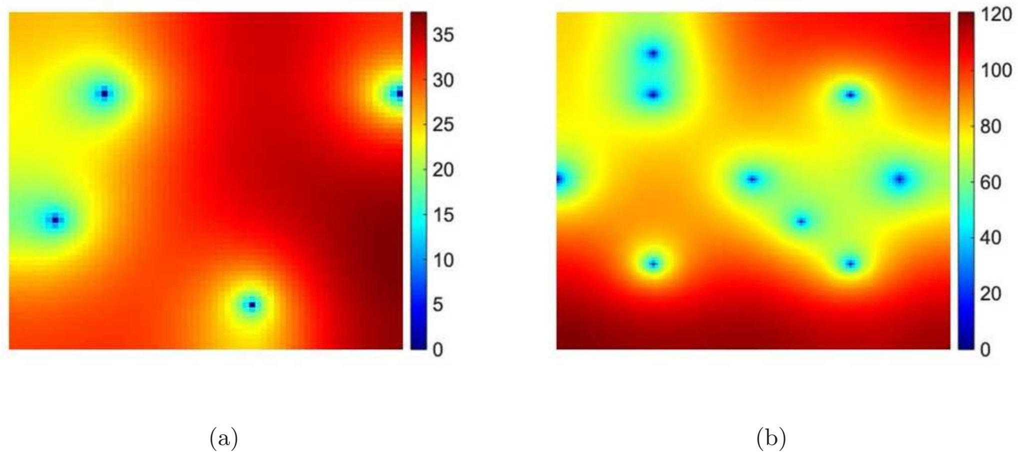
Figure 5.3: (a) Power and ground networks of Cortex-M0 DesignStart; (b) voltage drop map of the power network of (a).

power grid information including layout geometry, layer, via, as well as branch currents for later simulation.

Having a sufficient amount of training data is a crucial requirement for machine learning approaches. The GAN-based EM-induced IR drop prediction requires a lot of power grid samples and their corresponding ground truth EM-induced IR drop along the aging time. However, synthesizing a large number of designs and dumping their power grid information is not realistic. We first synthesized three power grid designs, and then for each design, we generated 10000 different workloads respectively. The network samples have the same topology as the synthesized designs. Although they have the same number of power strips, they differ in the branch width and length, thus the wire can be sized properly later on for EM-induced IR drop fixing.

Data representation

Representation of data has a tremendous impact on GAN behavior. To preserve the geometric and spatial relationship, we first encode the power grid workloads and voltages into matrices and then convert the voltage matrix into red-green-blue (RGB) channels of images, as illustrated in Fig. 5.4. Each pixel stands for one node, the length and width information are discarded, while the relative position of each node and its voltage value are kept. Such compact representation will dramatically reduce the image size compared with the representation from Fig. 5.3(b), which will further speed up the training process.



(a) (b)

Figure 5.4: Compact IR drop image of power grid networks of (a) Design 2: 4k nodes; (b) Design 3: 16k nodes.

As the pixels in our images are not RGB colors but real voltage values instead, they usually do not change dramatically, e.g., the maximum voltage value is 1.05V and most values fall in the range [0.7 1.05]. The channels of input are real resistance and current, thus they have the same numerical problem. Such a small numerical range is not suitable

for neural networks. As a result, we rescaled all data in the training to the range between -1 and 1.

5.2.4 *GridNet* architecture

GAN is a neural network model widely used in unsupervised machine learning tasks. A traditional GAN is composed of two separate deep neural networks, one is generator G and the other is discriminator D, there is no control on modes of the data being generated. In the CGAN model, the generator learns to generate a fake sample with a specific condition rather than a generic sample from unknown noise distribution.

For our problem, *GridNet* does not generate voltage maps from the random noises, instead, the inputs are the selected electrical and implicit geometrical features of the power grid networks and aging time. In order to implicitly learn the distribution of the voltage and map it to the corresponding 2D voltage image, we use a CGAN as the backbone for our model shown in Fig. 5.5. As we can see, to make the GAN model learn the temporal dynamics of EM-induced IR drops, we use the time variable as the continuous condition for both generator and discriminator, which was demonstrated to be effective for financial market risk analysis [27].

Take a power grid design with 120 rows and 120 columns as an example, there are five channels of input for the generator: the column resistance image $img_{col} \in \mathbb{R}^{119\times120\times1}$, the row resistance image $img_{row} \in \mathbb{R}^{120\times119\times1}$, the current source image $img_{cur} \in \mathbb{R}^{120\times120\times1}$, wire length l, and aging time t. t and l are expanded into $\mathbb{R}^{128\times128\times1}$ by channel-wise duplication, respectively. In addition, the three images are all expanded to the same size, such

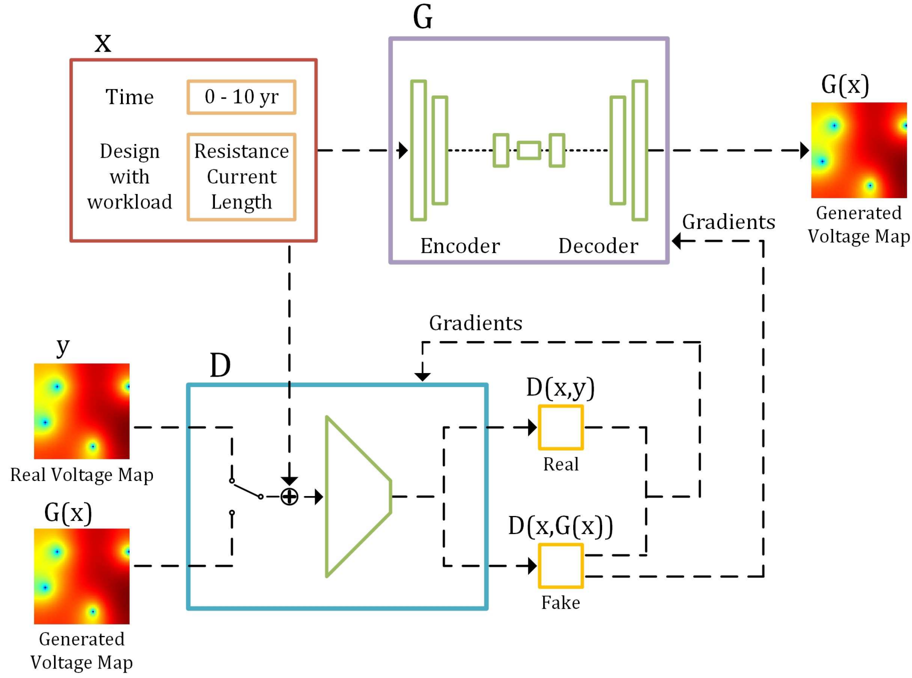


Figure 5.5: The CGAN architecture for EM-induced voltage prediction.

that img_{col}, img_{row}, img_{cur}, l, and t can be concatenated depth-wise. The resulted input x given to the generator is a $128 \times 128 \times 5$ tensor with all entries normalized as described in the previous section. We employ an encoder-decoder architecture as our generator that is widely used in image-to-image applications. The input is downsampled through a series of convolutional layers until a bottleneck layer, at which the latent features are extracted and then reversely upsampled through transposed convolutional layers. The generator is trained to extract useful latent features from the input and then reconstruct the output voltage map basing on this information.

The output of the generator is a voltage map, which is denoted as $G(x)$. Either the generated $G(x)$ or the real EM-induced voltage image y is fed into the discriminator D alternatively together with its corresponding workloads and aging time x as the condition

input. The output of the discriminator is denoted as $D(G(x), x)$ or $D(y, x)$ depending on whether the generated or the real EM-induced voltage image was inputted. In the training process, we use the Wasserstein Distance [6] as the measurement of the difference between the real and the generated EM-induced voltage image distribution to take advantage of higher stability and convergence possibility.

5.2.5 Fast sensitivity calculation using the automatic differentiation in DNNs

One important observation for all the deep neural networks including the GAN model is that they are all differentiable with respect to the model weights so that training can be performed by sensitivity/gradient information via the automatic differentiation scheme, specifically the back-propagation algorithm.

In this work, we leverage the existing automatic differentiation to compute the sensitivity information between the output and all of the input resistance through *GridNet*. To be specific, we can compute the partial derivatives of one output voltage map with respect to every input resistance in one back-propagation (same cost of one inference) of the generator DNN network using the Tensorflow tf.gradients API, which is exactly the same technique employed in the training process. The only difference here is that the derivative is taken with respect to the input of the generator instead of the trainable variables in the model. In other words, one has to perform one inference using *GridNet* to compute sensitivity for k resistances for one output node. Our sensitivity calculation is similar to the adjoint network-based approach [22], however, this method requires two simulations of the *EMspice* for the original and adjoint networks for each output node. In our case, we do

not require computing the sensitivities for all the output nodes, instead, we only focus on a few nodes that are subject to IR drop violations, which makes the sensitivity computation even more efficient.

5.3 Fast layout fixing for EM-induced IR failures

Power network is usually synthesized at an early stage of the chip design flow. Given a power gird, it is possible that it is vulnerable at the target time T_{target}. In other words, the maximum IR drop exceeds the voltage drop threshold $V_{drop_{th}}$, thus the design flaws need to be fixed. Specifically, the EM-lifetime of the power grids refers to the time at which an EM-induced voltage failure is expected to happen. There are two failure scenarios: vulnerable in the initial state; robust initially, but has EM-induced voltage violations at T_{target}.

In this section, we present two fast localized fixing methods based on *GridNet*. The first method uses a relatively rough estimation to expediently fix IR drop failures at the preroute stage. In contrast, the second method only takes the second failure scenario into account. It benefits from the gradient information obtained from *GridNet*. We assume that only a few EM-induced IR drop violations will occur at T_{target} during the later power grid design stages such as EM sign-off. This is typically the situation as the power grid network has been well designed at the synthesis step with EM failure considerations.

5.3.1 Fast localized IR drop fixing

The first localized fix method tries to size the whole interconnect tree one at a time until we meet the IR drop constraint at the targeted lifetime. Specifically, after a power grid network is synthesized, multiple voltage violations may occur at T_{target}. With *GridNet*, we can easily obtain all the nodal voltages at T_{target}. In addition, *GridNet* is able to provide the EM lifetime of the power grid.

Starting from the original power grids, widening one interconnect tree can have inevitable effects on nearby trees, namely, fixing several critical trees may be sufficient to fix voltage violations on all trees within this local area. Therefore, there is no need to widen all vulnerable trees at the same time, which would result in a large design overhead. In our method, if voltage violations happen, the interconnect with the largest IR drop will be widened by a scaling factor s, where $s > 1$ and it is determined by experiments.

According to Moudallal *et al.* [42], the voltage drop is a monotonically increasing function with respect to aging time given a certain grid sample, thus we only need to look at T_{target} without considering any time in between.

The modified power grid information is then fed into *GridNet* to get the newly predicted nodal voltages at T_{target}. We iteratively predict voltages and widen one tree at a time until the IR drops of all nodes are bounded within $V_{DD} - V_{drop_{th}}$.

According to the design rules and the minimum allowed space for standard cell placement and routing, the interconnect wire must be modified in a manner under some specific constraints. In our method, we specify that each wire can be widened under those design rules, but they have a maximum allowed value.

5.3.2 Sensitivity-based localized IR drop fixing

As we discussed, the sensitivity information of node voltage with respect to the wire resistance can be obtained as a by-product from the CGAN model of *GridNet* for each given input design using simple back-propagation as mentioned in Chapter 5.2.5. Typically the candidate wires (and their wire segments) are the wires with or close to void nucleations. As a result, those wires are good candidates for fixing as they are the ones causing the EM-induced IR drop failures.

Specifically, we assume that we have n violation nodes whose nodal voltages are represented by v_i, $i \in \{1, .., n\}$ and m wire segments whose with are represented by w_i, $i \in \{1, ..., m\}$. Then we can compute the following partial sensitivity matrix $\mathbf{S}_{n\times m}$:

$$\mathbf{S}_{n\times m} = \begin{bmatrix} \frac{\partial v_1}{\partial w_1} & \frac{\partial v_1}{\partial w_2} & \cdots & \frac{\partial v_1}{\partial w_m} \\ \frac{\partial v_2}{\partial w_1} & \frac{\partial v_2}{\partial w_2} & \cdots & \frac{\partial v_2}{\partial v_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_n}{\partial w_1} & \frac{\partial v_n}{\partial w_2} & \cdots & \frac{\partial v_n}{\partial w_m} \end{bmatrix} \tag{5.2}$$

Let $\Delta\mathbf{V} = [\Delta v_1, ..., \Delta v_n]$ represent the voltage drop changes we expected to meet IR drop constraint at T_{target} and $\Delta\mathbf{W} = [\Delta w_1, ..., \Delta w_m]$ be the required first-order width changes to make the voltage drop change. Then we have

$$\Delta\mathbf{V} = \mathbf{S} \times \Delta\mathbf{W} \tag{5.3}$$

To solve $\Delta\mathbf{W}$, we perform the least square regression as follows:

$$\begin{aligned} \mathbf{S}^T\Delta\mathbf{V} &= \mathbf{S}^T\mathbf{S} \times \Delta\mathbf{W} \\ \Delta\mathbf{W} &= (\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T\Delta\mathbf{V} \end{aligned} \tag{5.4}$$

We note that the required width changes $\Delta \mathbf{W}$ may also be subject to the design rules as there is an upper bound for the width. Once we update the width changes to the power grid network, we will run *GridNet* to verify IR drop at T_{target}. If there are still IR drop violations, more sensitivity-based fixing will be performed until there is no IR drop violation. The final design will be validated by *EMspice*.

5.4 Experimental results and discussions

5.4.1 Experiment setup

The above EM-induced IR drop prediction model for power grids (*GridNet*) has been implemented in Python using the *TensorFlow* library. The voltage violation fixing methods are implemented in Python. The experiments were carried out on a Linux server with 2 Xeon E5-2698v2 2.3GHz processors and Nvidia Titan X GPU.

In order to validate our work, we start from the power grid of the Cortex-M0 DesignStart processor. It is a 32-bit processor that implements the ARMv6-M architecture. This processor is synthesized using Synopsys Design Compiler and is placed and routed with Synopsys 32/28nm Generic Library. The power grid of Cortex has two layers, and there are 1k nodes in total.

Power grid information obtained from Synopsys IC Compiler is then fed into the power grid parser. The information includes but is not limited to structure, node location, wire layer, wire length, current source, voltage source and resistance values. The netlist format extracted from the grids agrees with IBM power grid benchmarks [43]. In order to

obtain enough power grids for training, we generate lots of synthesized IBM-format power grid networks so that different workloads can be tested and verified.

We train our DNN model using three different designs/topologies and each of them has a dataset containing 12000 pairs of (workloads and aging time, EM-induced IR drop) samples. *Design 1* comes from Cortex-M0, *Design 2* and *Design 3* are shown in Fig. 5.4(a) and Fig. 5.4(b), respectively. The temperature used in the experiment is 373K, the IR drop threshold $V_{drop_{th}}$ is $10\% V_{DD}$ and the target EM lifetime T_{target} is set to 10 years. For each workload, we collect the EM-induced IR drop results obtained by *EMspice* at 11 discrete aging time instants (0 to 10 years). We randomly select 15% of workloads for testing and the remaining 85% are assigned for the training set. During the training phase, all samples are randomly permuted at the beginning of every epoch.

5.4.2 EM-induced IR drop prediction results

Accuracy analysis

Once the *GridNet* model is trained, the generator is preserved and serves as the model for inference. The model can take any power grid workload with the same topology as input and give the predicted EM-induced voltage at a specified aging year. The predicted results from *GridNet* are compared with the baseline, which are the simulation results from *EMspice*. To evaluate the estimation error, we employ the root-mean-square error (RMSE) and as the metrics

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y' - y)^2}{N}} \tag{5.5}$$

where y' and y are the predicted and real voltage value, respectively. N is the total number of nodes. We evaluate our trained *GridNet* model on the testing set which was set aside during the training phase. The workloads in the testing set were randomly generated in the same way as the training set was produced. The random generation process guarantees that there is no overlap between these two datasets. The details and results are shown in Table. 5.1.

Table 5.1: Prediction results of different designs.

circuit	# nodes	# voltage sources	V_{DD} (V)	RMSE (mV)
Design 1	1024	2	1.05	5.697
Design 2	4096	4	1.05	6.100
Design 3	16384	9	1.05	3.922

A total number of 1800 different workloads are tested for each design. For each workload, 11 voltage images at 0 to 10 discrete aging years are generated. As can be seen from Table 5.1, comparing all 19800 generated EM-induced voltage images with the baseline on *Design 1*, *GridNet* achieves an average RMSE of 5.697mV, which represents about 0.57% error for a 1.05V power supply.

We randomly pick one testing workload from *Design 2* and compare the EM-induced voltage estimation at different aging years with the baseline in Fig. 5.6. Fig. 5.6(a) shows the predicted and real voltages at 0, 6 and 10 years, respectively. Since there is no obvious difference that can be seen, we zoom in the upper right corner of the design and use Fig. 5.6(c) to demonstrate the prediction error. Initially, the error distributes evenly, while at the end of the 6th year the error grows larger at two spots. The maximum error at 10 years is only 0.04mV, which indicates a good prediction. Fig. 5.6(b) illustrates the

EM-induced IR drop increasing process. The left figure shows that from 0 to 6 years, the IR drop at some spots increases faster than the other nodes. When comparing the right figure to the left, we can find that the spots are spreading over time. The reason is that there are nucleated voids nearby and the EM aging process leads to the resistance increase.

We further compare the predicted EM-induced IR drop with the baseline on *Design 1.1* and *Design 1.2*. To validate whether our model can successfully predict the IR drop of unseen power grids, all the testing data is separated from training data. The error statistics are shown in Table. 5.2, where *std* stands for standard deviation.

Table 5.2: Error statistics of Design 1.1 and Design 1.2.

		Design 1.1	Design 1.2
initially	mean (μV)	0.8165	18.8496
	max (mV)	3.4879	3.8722
	std (μV)	0.6115	0.9303
	RMSE (mV)	1.0199	2.1058
at 3rd aging year	mean (μV)	0.7382	31.1950
	max (mV)	2.5299	12.8284
	std (μV)	0.6999	1.2347
	RMSE (mV)	1.0170	3.3547
at 10th aging year	mean (μV)	-0.3939	82.7170
	max (mV)	1.1503	27.0619
	std (μV)	0.8030	2.5941
	RMSE (mV)	0.8941	8.6685

Design 1.1 is a power grid with one mortal interconnect and the initial maximum IR drop is 58.75mV. After one year, the resistance of the mortal interconnect begins to increase due to EM and the value is changing over time. After 10 years, the maximum IR drop becomes 59.54mV, which means the EM lifetime meets the 10-year target. In contrast, the predicted IR drop in the initial state and after 10 years are 57.93mV and 59.95mV,

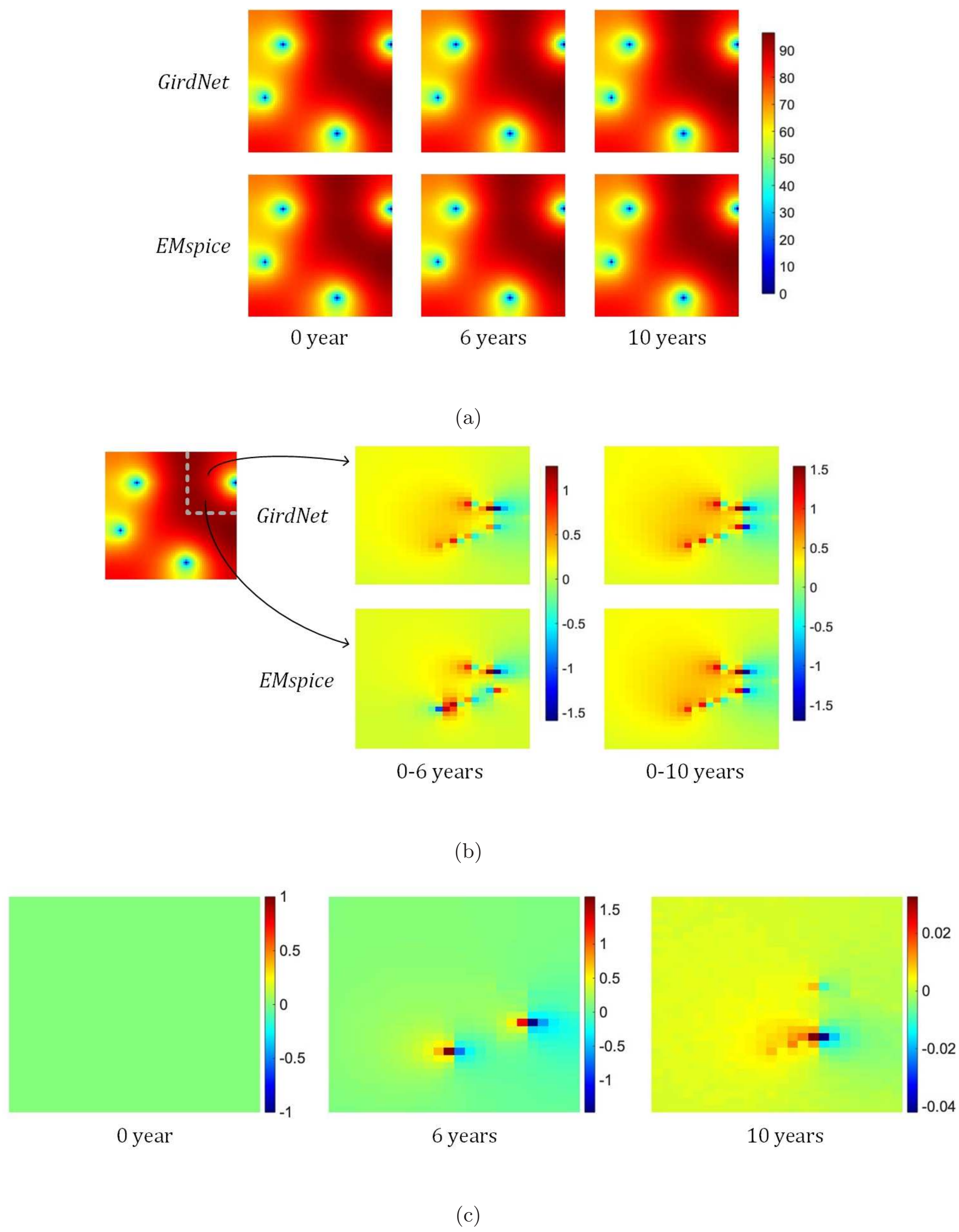


Figure 5.6: Comparison of inference results from *GridNet* and *EMspice*: (a) IR drop distribution at different years; (b) increased IR drop due to EM; (c) predicted voltage error.

respectively. Fig. 5.7(a) presents the correlation between the predicted EM-induced IR drop and baseline from 0 year to 10 years, with one-year interval, e.g., the purple dots indicate IR drop at 10 years. The errors of all predicted values are less than 3.5mV. The average error is 4.04×10^{-4}mV, with a standard deviation of 7.52×10^{-4}mV.

Design 1.2 is a power grid with 6 mortal interconnects and its EM-lifetime is just 3 years. Initially, the real maximum IR drop is 84.47mV whereas the predicted maximum IR drop is 82.34mV. After 3 years, the EM-induced values become 84.58mV and 85.56mV for the baseline and predicted values respectively. From the 4th year, wire resistance starts to increase, which will have large impacts on the whole grid. As a result, both the baseline and predicted one have the maximum IR drop larger than 110.83mV, resulting in a power grid failure. Finally, in the 10th year, the baseline and the predicted IR drop value are 133.99mV and 127.39mV, respectively. From Fig. 5.7(a) and Fig. 5.7(b), we can see that the correlations for different years in the first figure have similar patterns. On the other hand, the second figure looks different, the data for the first few years are concentrated in the lower part and the data for the last few years are distributed throughout the whole figure. The reason is that the EM effect is more clearly reflected in *Design 1.2*, which has a larger resistance increase.

The error histogram in Fig. 5.8 shows the error distribution of the predicted results. In Fig. 5.8(a), most of the predicted results are concentrated near the center, which agrees with the above analysis. As for Fig. 5.8(b), the error distribution for different years varies, because there are relatively large changes in the nodal voltages due to EM. When we look at

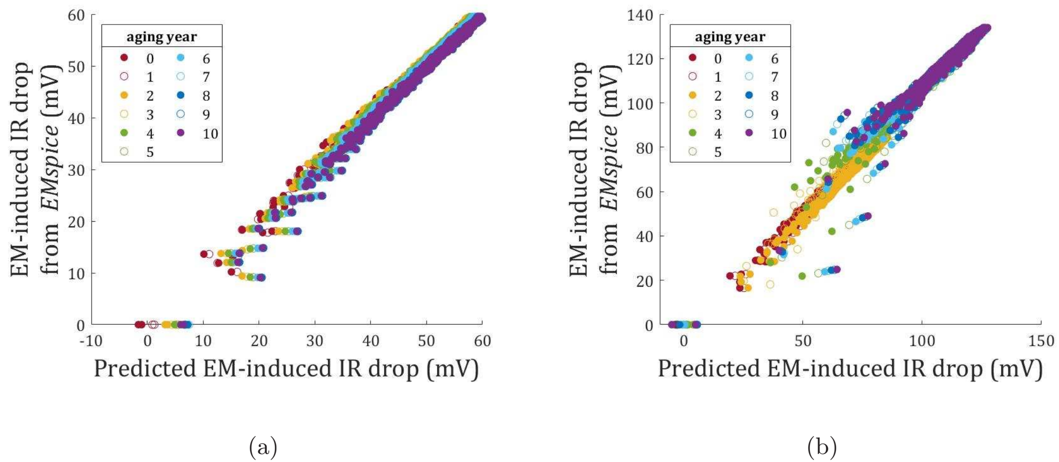


Figure 5.7: Predicted IR drops versus the baseline of (a) Design 1.1; (b) Design 1.2.

the details of each year, we can find that the error still follows the uniform distribution. For instance, in the 10th year, 87.79% of the nodes have errors between 8.13mV and 13.8mV.

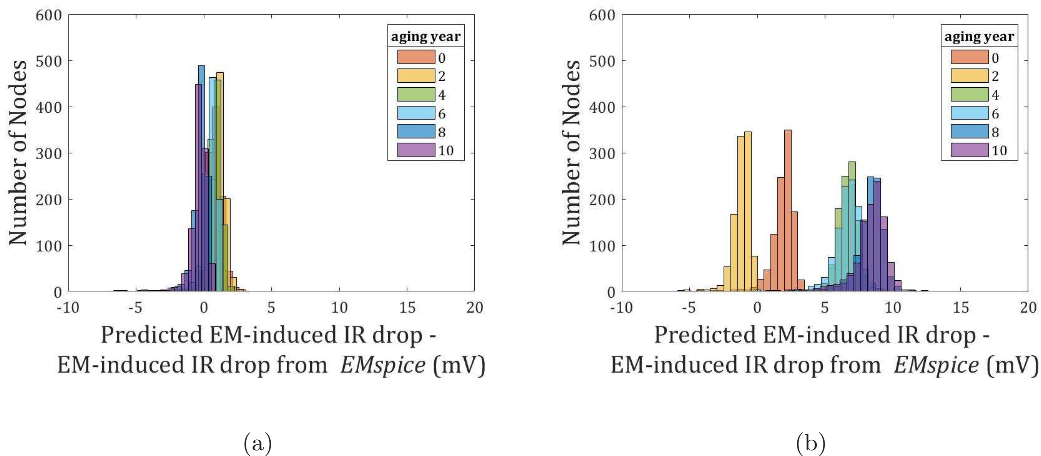


Figure 5.8: Error histogram of (a) Design 1.1; (b) Design 1.2.

We further do an accuracy study on the EM lifetime prediction for power grid networks. We randomly select 700 mortal designs/workloads from the testing sets, the lifetime prediction results are shown in Fig. 5.9. Among the designs, the prediction from

GridNet is consistent with the baseline for 83% of cases. If we allow two-year prediction errors, then it agrees with the baseline in 90.86% of cases. The accuracy degradation is due to the fact that lifetime is more related to the minimum voltage rather than all nodal voltages.

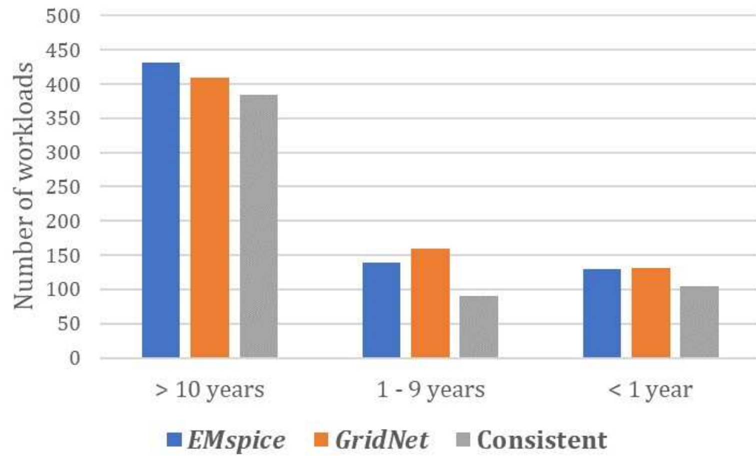


Figure 5.9: Predicted lifetime versus baseline.

Speedup analysis

In what follows, we provide a comparison of speed between *GridNet* and the baseline *EMspice* on EM-induced voltage analysis. We randomly pulled the designs from both training and testing set from *Design 1*, which contains 1k nodes. Both our *GridNet* model and *EMspice* were tested to generate the voltages from the initial state until 10 years. Specifically, the simulation time step of *EMspice* is set to one month. The experiments were performed on the same server and the accumulating time costs on 500 designs are plotted in Fig. 5.10.

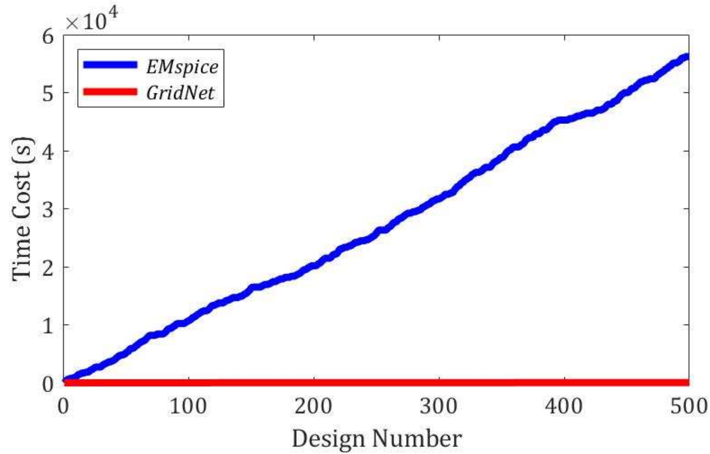


Figure 5.10: Prediction and simulation speed comparison between *GridNet* and *EMspice*.

The total computing time on the 500 designs is 31.26h and 10.0s for *EMspice* and *GridNet*, respectively, indicating that about 11232 or $10^4\times$ speedup over *EMspice*. For *EMspice*, the time cost on the estimation of a single design varies from 0.57s to 427s depending on the EM immortality condition. For *GridNet*, however, the inference speed is steadily around 5ms for all the designs. The computing cost of *GridNet* is invariant to immortality conditions, which makes it much more suitable for larger-scale designs and leads to better scalability.

As for larger designs, the speedup becomes even more significant since the simulation time for *EMspice* grows considerably. For instance, for *Design 3* which has 16k nodes, obtaining the EM-induced IR drop result at the 10th aging year takes more than 1.5h. If applying *GridNet*, the inference time will be around 10ms, which indicates that the speedup will be more than 5×10^5.

5.4.3 Fast IR drop fixing results

The results of the fast IR drop fixing method using prediction results from *GridNet* are listed in Table 5.3. The last column in this table and following table indicate CPU times used for the fixing process including data processing and inference costs of *GridNet*. There exist some benchmarks which do not satisfy the target EM lifetime with their original width configuration. In this case, we widened the most vulnerable interconnects to make the tree less likely to fail. As we can see, in *Design 1.a* example, there are 9 mortal wires in the beginning. At the same time, the minimum voltage of the power grid is 0.9201V, which is a 12.37% voltage drop, meaning that the EM lifetime is 0. Following the method discussed in Chapter 5.3.1, after the first time width adjustment, the number of mortal wires reduce from 9 to 7, however, the maximum voltage drop is 10.8% which still does not meet our design requirement. In the second iteration, we keep modifying the same interconnect, after that, the number of mortal trees is reduced to 6, with a 1.93% area increase compared to the original design. In the third iteration, another interconnect is widened, after modification, there still exist 5 mortal wires, however, all the voltage drops in the initial state are within 10%, and after 10 years, the minimum nodal voltage is 0.9461V, which meets the 10 year target.

Table 5.3: Fast IR drop fixing results.

circuit	# mortal wires	original lifetime	# iter	# widened wires	area increase	total time
Design 1.a	9	0 yr	3	2	3.31%	0.6644s
Design 1.b	3	8 yr	1	1	0.45%	0.2865s
Design 2.a	6	3 yr	2	2	1.75%	1.6461s
Design 2.b	4	6 yr	2	1	1.08%	1.3643s

5.4.4 Sensitivity-based localized fixing results

Table 5.4 shows results from the sensitivity-based localized fixing scheme. In *Design 1.c* example, the predicted lifetime of the power grid network is 9 years. This gird has 3 mortal wires initially and 2 predicted IR drop violations at T_{target}. After finding the violations and using the sensitivity information from *GridNet*, we find that all the 3 branches with void nucleations have to be widened. It turns out that with just one iteration, the modified network meets the 10-year lifetime target. This result is also verified by the simulation results from *EMspice*. *Design 1.d* is a power grid with 6 mortal wires and its predicted lifetime is 7 years. At T_{target}, there are 13 IR drop violations and the minimum voltage is 0.9435V. In the first iteration, 2 branches are widened according to Eq.(5.4). Afterward, the minimum voltage is increased to 0.9442V and the number of violations is reduced to 8. Then the second iteration is performed, after which only 3 violations are left. Finally, the third iteration leads to a minimum voltage of 0.9452V. During the iterations, 5 branches are widened in total.

Table 5.4: Sensitivity-based localized fixing results.

circuit	# mortal wires	# violations at T_{target}	# iter	# modified branches	area increase	total time
Design 1.c	3	2	1	3	0.446%	1.62s
Design 1.d	6	13	3	5	0.765%	3.91s
Design 2.c	4	3	1	2	0.151%	1.07s
Design 2.d	7	9	2	6	0.352%	1.86s

According to the results in Table 5.4, the sensitivity-based localized fixing method is very efficient in fixing the IR drop violations in a localized style. Since this method enables localized branch fixing rather than whole interconnect modification, it keeps most

branches unchanged, thus is more suitable to perform in the IR and EM sign-off stages of the physical design flow. Compared with Table 5.3, we can see that the sensitivity-based method leads to less area overhead compared to the first IR drop fixing method as we can select the most profitable segments to size.

On the other hand, due to the maximum allowable power routing space and other design rule constraints, IR drop reduction can be achieved by modifying only the branches with void nucleations or the branches close to the wires with void nucleations. As a result, a more expensive global wire fixing or optimization method may still be needed.

5.5 Summary

In this paper, we have presented *GridNet*, a fast data-driven EM-induced IR drop analysis framework for power grid networks based on the CGAN model. It is able to speed up the incremental full-chip EM-induced IR drop analysis in sensitivity-based optimization and IR drop violation fixing during the power grid design and optimization. We demonstrated that the *GridNet* can be adopted to learn temporal dynamics in the aging process of power grid networks by using the continuous time as one of the conditions. Numerical results on a number of synthesized power grid networks validated that the new method can lead to five orders of magnitude speedup over the recently proposed full-chip coupled EM and IR drop analysis tool. More importantly, by leveraging the differentiable feature of the *GridNet* model, we can easily obtain the sensitivity information of node voltage with respect to the wire resistance or width. We then demonstrated two efficient localized strategies to fix IR drop violations for late-stage power grid designs. Numerical results showed that the

localized IR drop violation fixing is remarkably fast by utilizing the sensitivity information from *GridNet*.

Chapter 6

GridNetOpt: Fast full-chip EM-aware IR drop constrained power grid optimization via deep neural networks

6.1 Related works

IR drop estimation problem can be accelerated by applying state-of-the-art machine learning techniques such as artificial neural network (ANN), XGBoost, and convolutional neural network (CNN). A good summary of recent developments for machine learning-based IR drop analysis can be found at [65]. Apart from IR drop estimation works, IR drop-related machine learning works have also been actively studied. Some works directly

estimate the impact of IR drop while some other works only view IR drop as one of the design constraints.

Recently, an EM-induced IR drop estimation tool *GridNet* has been proposed, which is based on the CGAN architecture to generate 2D IR drop maps from the netlist and electrical features [71]. It has been demonstrated that the trained CGAN models can be used to compute the sensitivity of objective function with respect to the wire resistances for localized power grid fixing. In this work, we further leverage *GridNet*-like DNN for full-chip power grid optimization subject to the EM-induced IR drop constraints.

6.2 New EM-induced voltage constrained optimization problem

6.2.1 Problem formulation

Let $G = \{N, B\}$ be a power grid network with n nodes $N = \{1, ..., n\}$ and b branches $B = \{1, ..., b\}$. Each branch i in B connects two nodes p and q with current flowing from p to q. l_i, w_i and g_i are the length, width and conductance of branch i, respectively. ρ is the sheet resistance. The width w_i of branch i is

$$w_i = \rho \frac{l_i}{r_i} = \rho l_i g_i \tag{6.1}$$

Objective function

We can express the total routing area of the power grid network in terms of sheet resistance, branch length, width and conductance as follows

$$a = \sum_{i \in B} l_i w_i = \sum_{i \in B} \rho l_i^2 g_i \tag{6.2}$$

The objective is to minimize the area of the network. Assume that the topology and physical locations of the network are fixed, ρl_i^2 will become a constant and can be expressed as α_i , then the objective function is simplified to bed

$$a = \sum_{i \in B} \alpha_i g_i \tag{6.3}$$

Constraints

The constraints that need to be satisfied for a reliable power grid network are shown as follows.

1. New EM-induced voltage drop constraints: When a void is nucleated and the interconnect enters into the growth phase, an increase over time in branch resistance will happen and may lead to time-varying nodal voltages.

   The voltage drop of the leaf node j at aging time t is limited by a constant

$$v_{dd} - v_{j,t} \leq u \tag{6.4}$$

   where v_{dd} is the supply voltage and u is the bound of the IR drop.

2. Minimum width constraints: Usually different layers have different requirements for the width of the metal wires

$$w_i \geq w_{i,min} \tag{6.5}$$

where $w_{i,min}$ is the minimal metal line width.

According to Eq. (6.1), the above equation can be rewritten as

$$g_i \geq \frac{w_{min}}{\rho l_i} \tag{6.6}$$

3. Kirchhoff's current law: Kirchhoff's current law defines the function of nodal voltages. In our approach, we view node voltages as functions of conductance, so it is implicitly satisfied.

6.2.2 Penalty method

The power grid optimization aims to minimize objective function (6.3) subject to constraints (6.4) and (6.5). It will be referred as problem P. Problem P is a constrained nonlinear optimization problem.

Penalty method is adopted to solve problem P. By adding a penalty term to the objective function that prescribes a high cost for the constraint violations, the original constrained problem is approximated with a sequence of unconstrained problems [64].

Penalty function formulation

We adopt a penalty function as follows:

$$f = a + p_t = a + \beta \cdot \sum_j c_{j,t}^2 \tag{6.7}$$

where a is the network area of function (6.3), p_t is the penalty term and β is the penalty parameter. For the voltage drop constraint violation

$$c_{j,t} = \begin{cases} 0, & \text{if } v_{j,t} \geq v_{dd} - u \\ v_{j,t} - (v_{dd} - u), & \text{else} \end{cases} \tag{6.8}$$

Given branch i, we simplify Eq. (6.8) as

$$c_{j,t} = v_{j,t} - (v_{dd} - u)\,, \text{for all } j \in E_{vdrop} \tag{6.9}$$

Minimal width constraints are not added into the penalty function (6.7), the reason is that this algorithm simply sets the branches that do not satisfy minimal width constraints with the minimal metal line width. The original constrained problem P is transformed to the problem of minimizing the penalty function (6.7) with minimal width constraints (6.5).

We first analyze the network for the node voltages and the branch currents while considering its aging time t and then identify the constraint violations. The conjugate gradient method is adopted to update branch widths during each iteration. The process stops when all the constraints are satisfied.

Moudallal *et al.* [42] observed that the IR drop $v_{dd} - v_{j,t}$ is a monotonically increasing function with respect to time, in other words, $v_{j,t_1} \geq v_{j,t_2}$ for $0 \leq t_1 \leq t_2$. Although branch resistance increase does not necessarily lead to an IR drop increase, this assumption holds in most cases. With this, we restrict our attention to the target aging time T, then Eq. (6.9) becomes

$$c_{j,T} = v_{j,T} - (v_{dd} - u)\,, \text{for all } j \in E_{vdrop} \tag{6.10}$$

Optimization scheme

Given an initial penalty parameter β, we minimize the penalty function. We then increase the value of the penalty parameter β for the next minimization iteration. The process continues until all the constraints are satisfied. In this way, the original prob-

lem is transformed into a sequence of unconstrained minimization problems. The solution procedure can be described as follows.

1. Set an initial value of penalty parameter β; initial conductance vector $G^{(0)}$; and error bound $\varepsilon_b > 0$.
2. Solve the unconstrained minimization problem (6.7), obtain the current conductance vector $G^{(k)}$.
3. If $p_t^{(k)} < \varepsilon_b$, then stop; else, update penalty parameter β, set $k = k + 1$, and go to step 2).

In practice, the selection of penalty parameter β proves to be tricky and difficult.

6.3 Optimization strategies

6.3.1 Conjugate gradient method

In the penalty method, the efficiency of solving unconstrained minimization dominates the execution time. The conjugate gradient method, which is a method between the steepest descent method and the Newton method, deflects the direction of the steepest descent method by adding to it a positive multiple of the direction used in the last step. This method only requires the first-order derivatives but overcomes the steepest descent method's shortcoming of slow convergence. At the same time, the method does not need to save and compute the second-order derivatives that are needed by the Newton method.

We notice that the conjugate gradient method has been used for IR drop constrained power distribution network optimization [64] and for on-chip power/ground network optimization considering decap as well [26]. The IR drop constrained work [64] shows

that the conjugate gradient-based optimization method is more scalable than the linear programming-based method [58]. However, this method is still based on the Black's based EM model, which adds limited current density constraint to each wire segment. It cannot optimize the power grids with nucleated wires whose target is a given lifetime. In our problem, we solve the EM-induced IR drop optimization problem at the target lifetime considering more complicated physics-based EM models. Specifically, the models involve many computationally intensive circuit-level EM-IR multi-physics simulations for full-chip power grid networks.

Flether-Reeves (F-R) conjugate gradient method is employed in this work. The steps are summarized as Algorithm 6.

Algorithm 6 Unconstrained power grid area minimization algorithm

---

Input: Current conductance vector G.

Output: New conductance vector G.

1: $k := 0$.

2: Set initial descent direction to negative direction of the gradient $P^{(k)} = -\nabla f\left(G^{(k)}\right)$.

3: /*F-R conjugate gradient method*/

4: **repeat**

5: Line search to determine a nonnegative scalar $\lambda_{opt}^{(k)}$ that minimizes f.

6: Update conductance vector $G^{(k+1)} = G^{(k)} + \lambda_k P^{(k)}$.

7: Choose new descent direction $P^{(k+1)}$.

8: $k := k + 1$.

9: **until** $\left\|\nabla f\left(G^{(k)}\right)\right\| < \varepsilon_{FR}$

6.3.2 DNN-based fast EM-induced IR drop estimation

The conjugate gradient optimization framework requires the sensitivity of panelized objective with respect to wire conductance or width. As we show later, this actually requires intensive full-chip coupled EM and voltage (IR drop) analysis using *EMspice*. Such circuit-level multi-physics-based full-chip power grid simulations are very expensive and even prohibitive for large problem sizes.

In this work, we build the learning-based models based on the physics-based simulation tool to accelerate the sensitivity calculation. Among different DNN candidates, we select GAN [28, 21] as the IR drop map analysis can be treated as a 2D image generation process from some given conditions or features. Specifically, conditional GAN is adopted to estimate EM-induced voltage maps via an unsupervised learning process based on the simulated data from *EMspice*. This CGAN model does not generate voltage maps from the random noises, instead, the inputs are the selected electrical and implicit geometrical features of the power grid networks and aging time, which are treated as conditions. In order to implicitly learn the distribution of the voltage and map it to the corresponding 2D voltage image, as stated in Chapter 5.2.4, CGAN architecture in Fig. 5.5 is applied as the backbone for our problem. As we can see, to make the GAN model learn the temporal dynamics of EM-induced IR drops, aging time t is used as the continuous condition for both generator and discriminator.

6.3.3 Gradient calculation for the objective function

In the first step of the F-R conjugate gradient method, we analyze the network and derive the node voltage and current flow. From Eq. (6.7), the partial derivative of penalty function with respect to branch conductance can be expressed as

$$\frac{\partial f}{\partial g_i} = \frac{\partial a}{\partial g_i} + \frac{\partial p_t}{\partial g_i} \tag{6.11}$$

The first term of Eq. (6.11) is equal to a constant α_i and the second term can be expanded easily. To sum up,

$$\frac{\partial f}{\partial g_i} = \alpha_i + \beta \cdot \sum_j \frac{\partial v_{j,t}}{\partial g_i} \cdot 2 \cdot c_{j,t}, \text{for all } j \in E_{vdrop} \tag{6.12}$$

Since our main focus is to ensure that the EM-induced voltages at target time T do not have violations, it is enough to search for a solution that decreases voltage drops at time T.

$$\frac{\partial f}{\partial g_i} = \alpha_i + \beta \cdot \sum_j \frac{\partial v_{j,T}}{\partial g_i} \cdot 2 \cdot c_{j,T}, \text{for all } j \in E_{vdrop} \tag{6.13}$$

Thus, the gradient of penalty function f with respect to conductance vector G is

$$\nabla f\,(G) = \left[\frac{\partial f}{\partial g_1}, \frac{\partial f}{\partial g_2}, \ldots, \frac{\partial f}{\partial g_i}, \ldots, \frac{\partial f}{\partial g_b}\right]^T \tag{6.14}$$

6.3.4 Gradient calculation via adjoint network method

Traditionally, the adjoint network method has been proposed to calculate the partial derivative of the node voltages with respect to branch conductance [22]. The adjoint network method can compute the sensitivity of one node voltage with respect to all resistance or conductance, but computing the sensitivities for all the node voltages will have very low efficiency. Instead of solving all adjoint networks separately, the merged adjoint

network method only needs to solve circuit equations twice to calculate the final gradient of the objective function [26]: one is for the original network and the other is for the merged adjoint network. In this work, we implement the merged adjoint network method for performance comparison.

Compute merged adjoint network using *EMspice*

Let N and $N'(j)$ be the original network and the adjoint network, respectively. The two networks have the same topology and conductance values. By running *EMspice* simulator, we can easily obtain conductance matrices of N and $N'(j)$ at time T. The only difference between the two networks is that all the absorbing current of $N'(j)$ is set to zero except node j. Since *EMspice* also tells the nodal voltages for N at time T, we only have to build $B(j)$ to solve the branch voltages for $N'(j)$.

$$B(j) = [0, 0, \ldots, -1, 0, \ldots, 0]^T \tag{6.15}$$

where -1 appears at index j.

Let $v_{i,T}$ and $v'_{i,T}$ denote branch i's voltage of N and $N'(j)$, the partial derivative of node voltage v_j with respect to the conductance of branch i is computed by

$$\frac{\partial v_{j,T}}{\partial g_i} = v_{i,T} \times v'_{i,T} = (v_{p,T} - v_{q,T}) \times \left(v'_{p,T}(j) - v'_{q,T}(j)\right) \tag{6.16}$$

Then Eq. (6.13) becomes

$$\frac{\partial f}{\partial g_i} = \alpha_i + 2 \cdot \beta \cdot (v_{p,T} - v_{q,T}) \times \left(\sum_j v'_{p,T}(j)\, c_{j,T} - \sum_j v'_{q,T}(j)\, c_{j,T} \right) \tag{6.17}$$

Suppose $V'(j)$ is a vector formed by the nodal voltages of $N'(j)$, we can have

$$v'_p(j) = C_p V'(j)\,, v'_q(j) = C_q V'(j) \tag{6.18}$$

where $C_p = [0, 0, \ldots, 0, 1, 0, \ldots, 0]$ with 1 appears at index p, and $C_q = [0, 0, \ldots, 0, 1, 0, \ldots, 0]$ with 1 appears at index q.

Therefore, Eq. (6.17) can be rewritten as

$$\frac{\partial f}{\partial g_i} = \alpha_i + 2 \cdot \beta \cdot (v_{p,T} - v_{q,T}) (C_p - C_q) \left(\sum_j c_{j,T} V'(j) \right) \tag{6.19}$$

6.3.5 Fast gradient calculation via deep neural networks

As mentioned earlier, sensitivity computation by adjoint network methods based on the detailed multi-physics *EMspice* simulation is computationally expensive. To mitigate this issue, we take advantage of the DNN-based model for sensitivity computation.

The objective of problem P is to minimize the power grid area while ensuring that the functional modules work properly at the target EM aging time T. Note that Eq. (6.1) holds only before the interconnect enters into the growth phase. Once the growth phase starts, the resistance starts increasing as current starts to flow through the more resistive barrier layer of the wire. In other words, the decrease in conductance g_i does not have an impact on the wire width w_i.

Let us add subscript time t to illustrate. Back to our EM-induced voltage drop constrained problem, the sensitivity value s we expect is $\partial v_{j,T}/\partial w_i$, which means the partial derivative of the node voltages at aging time T with respect to the branch width. According to Eq. (6.13), what we need to calculate is $\partial v_{j,T}/\partial g_i$. Since the width does not change during the EM process, i.e., $w_{i,T} = w_{i,0}$, it indicates that g_i here should be $g_{i,0}$. The rationale behind this is that we have to update conductance matrix $G^{(k)}$ for the next

iteration, however, updating $G^{(k)}$ implies updating width $W^{(k)}$, only the initial $W^{(k)}$ can be modified.

In *EMspice*, the coupled EM and IR simulation undergo a complex stress evolution and the change of EM-induced voltage drop with respect to time is nonlinear. From initial time 0 to target time T, the resistance of branch i may increase or remain unchanged, while the width of branch i always unchanged. It is impossible to express those partial derivatives using equations. Therefore, by applying the merged adjoint network method, we can easily get $\partial v_{j,t}/\partial g_{i,t}$, but cannot obtain $\partial v_{j,T}/\partial g_{i,0}$.

One important observation for all the deep neural networks is that they are all differentiable with respect to the model weights as well as data input as well so that training can be performed by sensitivity or gradient information via the automatic differentiation scheme, specifically the back-propagation algorithm.

In this work, we leverage the existing automatic differentiation scheme in DNN to compute the sensitivity information between the output and all of the input conductance through the CGAN model introduced in Chapter 6.3.2. To be specific, we can compute the partial derivatives of one output voltage map with respect to every input conductance in one back-propagation (same cost of one inference) of the generator DNN network using *TensorFlow*, which is exactly the same technique employed in the training process. The only difference is that the derivative is taken with respect to the input of the generator instead of the trainable parameters in the model. In other words, one has to perform one inference using the CGAN model to compute sensitivity for all b resistances for one output node. Specifically, we assume that we have m violation nodes at time T whose nodal voltages are

represented by v_j, $i \in \{1, .., m\}$. The CGAN model is able to give the estimated sensitivity values in milliseconds. Then we can compute the following partial sensitivity matrix $\mathbf{S}_{m\times b}$ easily

$$\mathbf{S}_{m\times b} = \begin{bmatrix} \frac{\partial v_{1,T}}{\partial g_{1,0}} & \frac{\partial v_{1,T}}{\partial g_{2,0}} & \cdots & \frac{\partial v_{1,T}}{\partial g_{b,0}} \\ \frac{\partial v_{2,T}}{\partial g_{1,0}} & \frac{\partial v_{2,T}}{\partial g_{2,0}} & \cdots & \frac{\partial v_{2,T}}{\partial g_{b,0}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_{m,T}}{\partial g_{1,0}} & \frac{\partial v_{m,T}}{\partial g_{2,0}} & \cdots & \frac{\partial v_{m,T}}{\partial g_{b,0}} \end{bmatrix} \tag{6.20}$$

More importantly, this automatic differentiation scheme is able to tell $\partial v_{j,T}/\partial r_{i,0}$ ($\partial v_{j,T}/\partial g_{i,0}$) directly, which is more reasonable to employ in our problem. With this, Eq. (6.13) via deep neural networks becomes

$$\frac{\partial f}{\partial g_i} = \alpha_i + \beta \cdot \sum_j \frac{\partial v_{j,T}}{\partial g_{i,0}} \cdot 2 \cdot c_{j,T}, \text{for all } j \in E_{vdrop} \tag{6.21}$$

6.4 Experimental results and discussions

The above EM-aware constrained power grid optimization (*GridNetOpt*) is implemented in Python with the *TensorFlow* library. The experiments are carried out on a Linux server with 2 Xeon E5-2698v2 2.3GHz processors and Nvidia Titan X RTX GPU with 24 GB memory.

In order to validate our work, we start from the power grid of the Cortex-M0 DesignStart processor. The power grid of Cortex has two layers, and there are 1k nodes in total. The netlist format extracted from the grids is consistent with IBM power grid benchmarks [43]. In order to obtain enough power grids with different EM conditions, we

generate lots of synthesized IBM-format power grid networks so that different workloads with different EM conditions can be tested and verified. In this work, we also tried an additional CNN model, and the results proved that the CGAN model can produce higher accuracy and more smooth node voltage images.

The maximum allowable IR drop is assumed to be 10% V_{dd} and the target EM aging time T is 10 years.

The power grids optimization results are shown in Table 6.1. *ARM-PG1* - *ARM-PG4* are the power grid network benchmarks. They have the same structure as the synthesized Cortex-M0 DesignStart processor, each of them has 64 interconnect wires and thousands of nodes but they differ in wire resistance, length, width and current sources. Therefore, the initial EM conditions of these benchmarks are different, such as the number of immortal wires. The number of violation nodes comes from the CGAN model for EM-induced voltage estimation.

Column 2 to 4 and 5 to 7 report the number of iterations (*# iter*), the reduced area ratio (*area reduced*) with respect to the original area and the total computational time (*time*) of our *GridNetOpt* and traditional adjoint network method with *EMspice*, respectively. From the results shown in the table, the area of both methods can be reduced after optimization and the area reduced ratio are similar. This demonstrates that our work achieves comparable optimization results compared to other conjugate gradient-based optimization works. We notice that for some cases the area reduced ratio is not big, the reason is that the test cases we used are already well-designed, the optimization space left is not large enough.

Table 6.1: Comparison of *GridNetOpt* against adjoint network method.

circuit	*GridNetOpt*			adjoint network method			speedup
	# iter	area reduced (%)	time (s)	# iter	area reduced (%)	time (s)	
ARM-PG1	4	0.29	54.79	4	2.56	624.61	11.40
ARM-PG2	6	0.69	38.45	8	0.25	787.81	20.49
ARM-PG3	6	1.42	35.34	5	0.18	631.49	17.87
ARM-PG4	6	0.76	45.08	3	0.12	662.34	14.69

With the *GridNetOpt*, we are able to meet the power grid lifetime target much faster than using the adjoint network method with *EMspice*, which would be a great advantage when the optimization space is not that large, because designers do not want to wait for a long time to only seek for a reduction potential. In the *ARM-PG2* example, the lifetime of the whole power grids is predicted less than 10 years because the maximum voltage at T exceeds $10\%V_{dd}$. There are 1 mortal wire and 84 violation nodes in total. Note that the lifetime definitions of individual interconnect wire and power grid network are different. The power grid lifetime refers to the earliest time t that EM-induced voltage violations of a power grid occurs, here, we do not care about the earliest time t, but cares about if there exist IR drop violations at target time T. By utilizing the by-product sensitivity information, we are able to get the optimization direction much easier as no complex calculation is required. Iteratively solving the unconstrained minimization problem and updating the conductance vector and penalty parameter, the power grid meets the lifetime target after 6 iterations. Even though *GridNetOpt* achieves better area reduction than using the adjoint network for this case, the iteration number of the former is fewer. There is no obvious relationship between reduced area and the number of iterations of the two methods that can be observed, e.g., *GridNetOpt* went through more iterations in the *ARM-PG1* case.

Consider both area reduction and computation time, *GridNetOpt* got similar area reduction but much better speedup (about **10x or more**) for all the cases, we can conclude that it outperforms the traditional adjoint network method for EM-induced IR drop constrained power grid optimization problem.

6.5 Summary

In this chapter, we presented a novel on-chip power distribution network wire sizing framework, called *GridNetOpt*, considering EM-induced IR drop constraints at the target aging time. The optimization framework is empowered by the data-driven learning-based time-varying IR drop modeling using deep neural networks. The new method can naturally leverage the differentiable feature of deep neural networks for fast sensitivity computation of node voltage with respect to wire resistance or width. Numerical results on a number of synthesized power grid benchmarks from ARM core CPU designs show that *GridNetOpt* can lead to an order of magnitude speedup over the conjugate gradient method with the traditional adjoint network approach.

Chapter 7

Conclusions

EM is a major failure effect for on-chip power grid networks of deep submicron VLSI circuits. EM degradation of metal grid lines can lead to excessive voltage drops before the target lifetime. This dissertation focuses on the presentation of new EM-aware power grid design and optimization techniques. In this chapter, the main contributions are summarized.

7.1 Summary of contributions

7.1.1 EM immortality and lifetime constrained power grid wire sizing

Chapter 3 presents a new power grid network sizing technique based on a fast voltage-based EM immortality check method for general multi-segment interconnect wires and a new physics-based EM assessment technique for more accurate time to failure analysis. We first show that the new power grid optimization problem, subject to the voltage IR drop and new EM constraints, can still be formulated as an efficient sequence of linear

programming problem. The new optimization will ensure that none of the wires fail if all the constraints are satisfied. However, requiring all the wires to be EM immortal can be over-constrained. To mitigate this problem, the first improvement is by means of adding reservoir branches to the mortal wires whose lifetime cannot be made immortal by wire sizing. This is a very effective approach as long as there is sufficient reservoir area. The second improvement is to consider the aging effects of interconnect wires in power grid networks. The idea is to allow some short-lived wires to fail and optimize the rest of the wires while considering the additional resistance caused by the failed wire segments. In this way, the resulting power grid networks can be optimized such that the target lifetime of the whole power grid networks can be ensured and they will become more robust and aging-aware over the expected lifetime of the chip. Numerical results on a number of IBM and self-generated power grid networks demonstrate that the new method can effectively reduce the area of the networks while ensuring immortality or enforcing target lifetime for all the wires, which is not the case for the existing current density constrained optimization methods.

7.1.2 Robust power grid network design considering EM aging effects

Chapter 4 presents a number of power grid network design and optimization techniques that consider the EM effects for multi-segment interconnect wires. First, we add a new void saturation volume-based EM immortality constraint, which helps reduce the conservativeness of the EM-aware on-chip power grid design. Along with the EM nucleation phase immortality constraint, we show that both EM immortality constraints can be naturally integrated into the existing programming-based power grid optimization framework.

Second, to mitigate the overly conservativeness of the immortality constrained optimization methods, we further explore three strategies: we first size up failed wires to meet one of the immorality conditions subject to the design rules; second, we consider the EM-induced aging effects on the power grid networks for a target lifetime, which allows some short-lived wires to fail and optimizes the rest of the wires; third, we present a large change sensitivity-based optimization scheme to perform localized fixing based on the recently proposed coupled EM-IR drop analysis method. Numerical results on a number of IBM-format power grid networks demonstrate that the new method can reduce more power grid area compared to the voltage-based EM immortality constrained optimizations. Moreover, the new method can optimize power grids with nucleated wires, which would not be possible with the existing methods. Results also show the sensitivity-based localized power girds fixing can fix EM-induced IR drop violations in a few minutes for synthesized power grid networks from ARM core designs.

7.1.3 Fast data-driven EM-induced IR drop prediction and localized fixing for on-chip power grid network

Chapter 5 presents a fast data-driven EM-induced IR drop analysis framework for power grid networks, named *GridNet*, based on the CGAN model. It aims to accelerate full-chip EM-induced IR drop analysis, as well as IR drop violation fixing during the power grid design and optimization. More importantly, *GridNet* can naturally leverage the differentiable feature of DNNs to obtain the sensitivity information of node voltage with respect to the wire resistance or width with marginal cost. *GridNet* treats continuous time and the given electrical features as input conditions, and the EM-induced time-varying voltage of

power grid networks as the conditional outputs, which are represented as data series images. We show that *GridNet* is able to learn the temporal dynamics of the aging process in the time domain. Besides, we can take advantage of the sensitivity information provided by *GridNet* to perform efficient localized IR drop violation fixing in the late-stage design and optimization. Numerical results on 36000 synthesized power grid network samples demonstrate that the new method can lead to $10^5\times$ speedup over the recently proposed full-chip coupled EM and IR drop analysis tool. We further show that localized IR drop violation fixing for the same set of power grid networks can be performed remarkably efficiently utilizing the cheap sensitivity computation from *GridNet*.

7.1.4 Fast full-chip EM-aware IR drop constrained power grid optimization via deep neural networks

Chapter 6 presents a fast full-chip EM-aware IR drop constrained optimization framework, named *GridNetOpt*, for on-chip power grid networks via a data-driven and learning-based IR drop estimation model. The new method can naturally take advantage of differentiable characteristics of DNNs for fast sensitivity computation of node voltage with respect to wire resistance or width. The method consists of several steps. First, a CGAN model is trained to obtain IR drops at a target aging time. The input power grids vary in wire sizes and current loads, and the training data comes from an advanced coupled EM-IR drop analysis tool. Second, the conjugate gradient-based optimization framework is employed. In the optimization flow, sensitivity information is computed from the CGAN model which is obtained trivially through its automatic differentiation operations for the power grid network candidates. Numerical results on a number of synthesized power grid

benchmarks from ARM core designs show that the *GridNetOpt* can lead to an order of magnitude speedup over the conjugate gradient-based method adopting the traditional adjoint network approach.

Bibliography

[1] International Technology Roadmap for Semiconductors 2.0 (ITRS 2.0), 2015. http://www.itrs2.net/itrs-reports.html.

[2] EMspice – Coupled EM-IR Analysis Tool for Full-Chip Power Grid EM Check and Sign-off, 2020. https://github.com/sheldonucr/EMspice.

[3] IEEE International Roadmap for Devices and Systems (IRDS), 2020. http://irds.ieee.org/.

[4] Syed M. Alam, Chee Lip Gan, Carl V. Thompson, and Donald E. Troxel. Reliability computer-aided design tool for full-chip electromigration analysis and comparison with different interconnect metallizations. *Microelectronics Journal*, 38(4-5):463–473, April-May 2007.

[5] Mohamed Baker Alawieh, Yibo Lin, Zaiwei Zhang, Meng Li, Qixing Huang, and David Z. Pan. GAN-SRAF: Sub-Resolution Assist Feature Generation Using Conditional Generative Adversarial Networks. In *Proceedings of the 56th Design Automation Conference*, DAC '19, pages 1–6, June 2019.

[6] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. *arXiv e-prints arXiv:1701.07875*, December 2017.

[7] H. B. Bakoglu. *Circuits, Interconnections, and Packaging for VLSI.* Addison-Wesley, Boston, MA, 1990.

[8] Mokhtar S. Bazaraa, Hanif D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms.* Wiley, Hoboken, NJ, 3rd edition, 2006.

[9] James R. Black. Electromigration-A Brief Survey and Some Recent Results. *IEEE Transactions on Electron Devices*, 16(4):338–347, April 1969.

[10] I. A. Blech. Electromigration in thin aluminum films on titanium nitride. *Journal of Applied Physics*, 47(4):1203–1208, April 1976.

[11] Norman Chang, Ajay Baranwal, Hao Zhuang, Ming-Chih Shih, Rahul Rajan, Yaowei Jia, Hui-Lun Liao, Ying-Shiun Li, Ting Ku, and Rex Lin. Machine Learning based

Generic Violation Waiver System with Application on Electromigration Sign-off. In *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*, ASP-DAC '18, pages 416–421, January 2018.

[12] Sandeep Chatterjee, Valeriy Sukharev, and Farid N. Najm. Fast Physics-Based Electromigration Checking for On-Die Power Grids. In *Proceedings of the 35th International Conference on Computer-Aided Design*, ICCAD '16, pages 1–8, November 2016.

[13] Sandeep Chatterjee, Valeriy Sukharev, and Farid N. Najm. Power Grid Electromigration Checking Using Physics-Based Models. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(7):1317–1330, July 2018.

[14] F. Chen, O. Bravo, K. Chanda, P. McLaughlin, T. Sullivan, J. Gill, J. Lloyd, R. Kontra, and J. Aitken. A Comprehensive Study of Low-k SiCOH TDDB Phenomena and Its Reliability Lifetime Model Development. In *Proceedings of the 44th International Reliability Physics Symposium*, IRPS '06, pages 46–53, March 2006.

[15] Hai-Bao Chen, Sheldon X.-D. Tan, Xin Huang, Taeyoung Kim, and Valeriy Sukharev. Analytical Modeling and Characterization of Electromigration Effects for Multibranch Interconnect Trees. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 35(11):1811–1824, November 2016.

[16] S. Chowdhury. Optimum Design of Reliable IC Power Networks Having General Graph Topologies. In *Proceedings of the 26th Design Automation Conference*, DAC '89, pages 787–790, June 1989.

[17] S. Chowdhury and Melvin A. Breuer. Optimum Design of IC Power/Ground Nets Subject to Reliability Constraints. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 7(7):787–796, July 1988.

[18] Salim U. Chowdhury and Melvin A. Breuer. Minimal Area Design of Power/Ground Nets Having Graph Topologies. *IEEE Transactions on Circuits and Systems*, 34(12):1441–1451, December 1987.

[19] Chase Cook, Zeyu Sun, Ertugrul Demircan, Mehul D. Shroff, and Sheldon X.-D. Tan. Fast Electromigration Stress Evolution Analysis for Interconnect Trees Using Krylov Subspace Method. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 26(5):969–980, May 2018.

[20] Chase Cook, Zeyu Sun, Taeyoung Kim, and Sheldon X.-D. Tan. Finite Difference Method for Electromigration Analysis of Multi-Branch Interconnects. In *Proceedings of the 13th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design*, SMACD '16, pages 1–4, June 2016.

[21] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1):53–65, January 2018.

[22] Stephen W. Director and Ronald A. Rohrer. The Generalized Adjoint Network and Network Sensitivities. *IEEE Transactions on Circuit Theory*, 16(3):318–323, August 1969.

[23] Dileep Divekar, Hal Daseking, and Ravi Apte. Fast DC Large Change Sensitivity Analysis for Circuit Simulation. In *Proceedings of the International Symposium on Circuits and Systems*, ISCAS '89, pages 2020–2022, May 1989.

[24] R. Dutta and M. Marek-Sadowska. Automatic Sizing of Power/Ground (P/G) Networks in VLSI. In *Proceedings of the 26th Design Automation Conference*, DAC '89, pages 783–786, June 1989.

[25] Yen-Chun Fang, Heng-Yi Lin, Min-Yan Sui, Chien-Mo Li, and Eric Jia-Wei Fang. Machine-learning-based Dynamic IR Drop Prediction for ECO. In *Proceedings of the 37th International Conference on Computer-Aided Design*, ICCAD '18, pages 1–7, November 2018.

[26] Jingjing Fu, Zuying Luo, Xianlong Hong, Yici Cai, Zhu Pan, and Sheldon X.-D. Tan. VLSI On-Chip Power/Ground Network Optimization Considering Decap Leakage Currents. In *Proceedings of the 10th Asia and South Pacific Design Automation Conference*, ASP-DAC '05, pages 735–738, January 2005.

[27] Rao Fu, Jie Chen, Shutian Zeng, Yiping Zhuang, and Agus Sudjianto. Time Series Simulation by Conditional Generative Adversarial Net. *arXiv e-prints arXiv:1904.11419*, April 2019.

[28] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. http://www.deeplearningbook.org.

[29] Stefan P. Hau-Riege and Carl V. Thompson. Experimental characterization and modeling of the reliability of interconnect trees. *Journal of Applied Physics*, 89(1):601–609, January 2001.

[30] Chia-Tung Ho and Andrew B. Kahng. IncPIRD: Fast Learning-Based Prediction of Incremental IR Drop. In *Proceedings of the 38th International Conference on Computer-Aided Design*, ICCAD '19, pages 1–8, November 2019.

[31] C.-K. Hu, D. Canaperi, S. T. Chen, L. M. Gignac, B. Herbst, S. Kaldor, M. Krishnan, E. Liniger, D. L. Rath, D. Restaino, R. Rosenberg, J. Rubino, S.-C. Seo, A. Simon, S. Smith, and W.-T. Tseng. Effects of overlayers on electromigration reliability improvement for Cu/low k interconnects. In *Proceedings of the 42nd International Reliability Physics Symposium*, IRPS '04, pages 222–228, April 2004.

[32] C.-K. Hu, M. B. Small, and P. S. Ho. Electromigration in Al(Cu) two-level structures: Effect of Cu and kinetics of damage formation. *Journal of Applied Physics*, 74(2):969–978, July 1993.

[33] Xin Huang, Armen Kteyan, Sheldon X.-D. Tan, and Valeriy Sukharev. Physics-Based Electromigration Models and Full-Chip Assessment for Power Grid Networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 35(11):1848–1861, November 2016.

[34] Xin Huang, Tan Yu, Valeriy Sukharev, and Sheldon X.-D. Tan. Physics-based Electromigration Assessment for Power Grid Networks. In *Proceedings of the 51st Design Automation Conference*, DAC '14, pages 1–6, June 2014.

[35] Renatas Jakushokas, Mikhail Popovich, Andrey V. Mezhiba, Selcuk Kose, and Eby G. Friedman. *Power Distribution Networks with On-Chip Decoupling Capacitors*. Springer, 2nd edition, 2011.

[36] M. A. Korhonen, P. Bo/rgesen, K. N. Tu, and Che-Yu Li. Stress evolution due to electromigration in confined metal lines. *Journal of Applied Physics*, 73(8):3790–3799, April 1993.

[37] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, May 2015.

[38] Jens Lienig and Matthias Thiele. *Fundamentals of Electromigration-Aware Integrated Circuit Design*. Springer, New York, NY, 2018.

[39] Shih-Yao Lin, Yen-Chun Fang, Yu-Ching Li, Yu-Cheng Liu, Tsung-Shan Yang, Shang-Chien Lin, Chien-Mo Li, and Eric Jia-Wei Fang. IR Drop Prediction of ECO-Revised Circuits Using Machine Learning. In *Proceedings of the 36th VLSI Test Symposium*, VTS '18, pages 1–6, April 2018.

[40] Jinglan Liu, Yukun Ding, Jianlei Yang, Ulf Schlichtmann, and Yiyu Shi. Generative Adversarial Network Based Scalable On-chip Noise Sensor Placement. In *Proceedings of the 30th International System-on-Chip Conference*, SOCC '17, pages 239–242, September 2017.

[41] Vivek Mishra and Sachin S. Sapatnekar. Predicting Electromigration Mortality Under Temperature and Product Lifetime Specifications. In *Proceedings of the 53rd Design Automation Conference*, DAC '16, pages 1–6, June 2016.

[42] Zahi Moudallal, Valeriy Sukharev, and Farid N. Najm. Power Grid Fixing for Electromigration-induced Voltage Failures. In *Proceedings of the 38th International Conference on Computer-Aided Design*, ICCAD '19, pages 1–8, November 2019.

[43] Sani R. Nassif. Power Grid Analysis Benchmarks. In *Proceedings of the 13th Asia and South Pacific Design Automation Conference*, ASP-DAC '08, pages 376–381, March 2008.

[44] Sachin S. Sapatnekar and Haihua Su. Analysis and Optimization of Power Grids. *IEEE Design & Test of Computers*, 20(3):7–15, May 2003.

[45] Valeriy Sukharev. Beyond Blacks equation: Full-chip EM/SM assessment in 3D IC stack. *Microelectronic Engineering*, 120:99–105, May 2014.

[46] Valeriy Sukharev, Xin Huang, Hai-Bao Chen, and Sheldon X.-D. Tan. IR-Drop Based Electromigration Assessment: Parametric Failure Chip-Scale Analysis. In *Proceedings of the 33rd International Conference on Computer-Aided Design*, ICCAD '14, pages 428–433, November 2014.

[47] Valeriy Sukharev, Armen Kteyan, Ehrenfried Zschech, and William D. Nix. Microstructure Effect on EM-Induced Degradations in Dual Inlaid Copper Interconnects. *IEEE Transactions on Device and Materials Reliability*, 9(1):87–97, March 2009.

[48] Valeriy Sukharev and Farid N. Najm. Electromigration Check: Where the Design and Reliability Methodologies Meet. *IEEE Transactions on Device and Materials Reliability*, 18(4):498–507, December 2018.

[49] Zeyu Sun, Ertugrul Demircan, Mehul D. Shroff, Chase Cook, and Sheldon X.-D. Tan. Fast Electromigration Immortality Analysis for Multisegment Copper Interconnect Wires. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(12):3137–3150, December 2018.

[50] Zeyu Sun, Ertugrul Demircan, Mehul D. Shroff, Taeyoung Kim, Xin Huang, and Sheldon X.-D. Tan. Voltage-Based Electromigration Immortality Check for General Multi-Branch Interconnects. In *Proceedings of the 35th International Conference on Computer-Aided Design*, ICCAD '16, pages 1–7, November 2016.

[51] Zeyu Sun, Sheriff Sadiqbatcha, Hengyang Zhao, and Sheldon X.-D. Tan. Accelerating Electromigration Aging for Fast Failure Detection for Nanometer ICs. In *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*, ASP-DAC '18, pages 623–630, January 2018.

[52] Zeyu Sun, Sheriff Sadiqbatcha, Hengyang Zhao, and Sheldon X.-D. Tan. Saturation-Volume Estimation for Multisegment Copper Interconnect Wires. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(7):1666–1674, July 2019.

[53] Zeyu Sun, Shuyuan Yu, Han Zhou, Yibo Liu, and Sheldon X.-D. Tan. EMSpice: Physics-Based Electromigration Check Using Coupled Electronic and Stress Simulation. *IEEE Transactions on Device and Materials Reliability*, 20(2):376–389, June 2020.

[54] Sheldon Tan, Mehdi Tahoori, Taeyoung Kim, Shengcheng Wang, and Zeyu Sun. *Long-Term Reliability of Nanometer VLSI Systems: Modeling, Analysis and Optimization.* Springer, New York, NY, 2019.

[55] Sheldon X.-D. Tan, Hussam Amrouch, Taeyoung Kim, Zeyu Sun, Chase Cook, and Joerg Henkel. Recent advances in EM and BTI induced reliability modeling, analysis and optimization. *Integration, the VLSI Journal*, 60:132–152, January 2018.

[56] Sheldon X.-D. Tan and C.-J. Richard Shi. Efficient Very Large Scale Integration Power/Ground Network Sizing Based on Equivalent Circuit Modeling. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(3):277–284, March 2003.

[57] Sheldon X.-D. Tan, C.-J. Richard Shi, and Jyh-Chwen Lee. Reliability-Constrained Area Optimization of VLSI Power/Ground Networks Via Sequence of Linear Programmings. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(12):1678–1684, December 2003.

[58] Xiang-Dong Tan, C.-J. Richard Shi, Dragos Lungeanu, Jyh-Chwen Lee, and Li-Pen Yuan. Reliability-Constrained Area Optimization of VLSI Power/Ground Networks Via Sequence of Linear Programmings. In *Proceedings of the 36th Design Automation Conference*, DAC '99, pages 78–83, June 1999.

[59] C. V. Thompson, S. P. Hau-Riege, and V. K. Andleigh. Modeling and experimental characterization of electromigration in interconnect trees. *AIP Conference Proceedings*, 491(1):62–73, 1999.

[60] A. H. Verbruggen, M. J. C. van den Homberg, L. C. Jacobs, A. J. Kalkman, J. R. Kraayeveld, and S. Radelaar. Resistance changes induced by the formation of a single void/hillock during electromigration. *AIP Conference Proceedings*, 418(1):135–146, 1998.

[61] Kai Wang and M. Marek-Sadowska. On-Chip Power-Supply Network Optimization Using Multigrid-Based Technique. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(3):407–417, March 2005.

[62] Xiaoyi Wang, Hongyu Wang, Jian He, Sheldon X.-D. Tan, Yici Cai, and Shengqi Yang. Physics-based Electromigration Modeling and Assessment for Multi-Segment Interconnects in Power Grid Networks. In *Proceedings of the 20th Design, Automation and Test in Europe*, DATE '17, pages 1727–1732, March 2017.

[63] Xiaoyi Wang, Yan Yan, Jian He, Sheldon X.-D. Tan, Chase Cook, and Shengqi Yang. Fast Physics-based Electromigration Analysis for Multi-Branch Interconnect Trees. In *Proceedings of the 36th International Conference on Computer-Aided Design*, ICCAD '17, pages 169–176, November 2017.

[64] Xiaohai Wu, Xianlong Hong, Yici Cai, Zuying Luo, Chung-Kuan Cheng, Jun Gu, and Wayne Dai. Area Minimization of Power Distribution Network Using Efficient Nonlinear Programming Techniques. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 23(7):1086–1094, July 2004.

[65] Zhiyao Xie, Hai Li, Xiaoqing Xu, Jiang Hu, and Yiran Chen. Fast IR Drop Estimation with Machine Learning. In *Proceedings of the 39th International Conference on Computer-Aided Design*, ICCAD '20, pages 1–8, November 2020.

[66] Zhiyao Xie, Haoxing Ren, Brucek Khailany, Ye Sheng, Santosh Santosh, Jiang Hu, and Yiran Chen. PowerNet: Transferable Dynamic IR Drop Estimation via Maximum Convolutional Neural Network. In *Proceedings of the 25th Asia and South Pacific Design Automation Conference*, ASP-DAC '20, pages 13–18, January 2020.

[67] Biying Xu, Yibo Lin, Xiyuan Tang, Shaolan Li, Linxiao Shen, Nan Sun, and David Z. Pan. WellGAN: Generative-Adversarial-Network-Guided Well Generation for Analog/Mixed-Signal Circuit Layout. In *Proceedings of the 56th Design Automation Conference*, DAC '19, pages 1–6, June 2019.

[68] Wei Ye, Mohamed Baker Alawieh, Yibo Lin, and David Z. Pan. LithoGAN: End-to-End Lithography Modeling with Generative Adversarial Networks. In *Proceedings of the 56th Design Automation Conference*, DAC '19, pages 1–6, June 2019.

[69] Lijuan Zhang. *Effects of scaling and grain structure on electromigration reliability of Cu interconnects*. PhD thesis, University of Texas, Austin, Austin, TX, 2010.

[70] Han Zhou, Liang Chen, and Sheldon X.-D. Tan. Robust Power Grid Network Design Considering EM Aging Effects for Multi-Segment Wires. *Integration, the VLSI Journal*, 77:38–47, March 2020.

[71] Han Zhou, Wentian Jin, and Sheldon X.-D. Tan. GridNet: Fast Data-Driven EM-Induced IR Drop Prediction and Localized Fixing for On-Chip Power Grid Networks. In *Proceedings of the 39th International Conference on Computer-Aided Design*, ICCAD '20, pages 1–9, November 2020.

[72] Han Zhou, Yijing Sun, Zeyu Sun, Hengyang Zhao, and Sheldon X.-D. Tan. Electromigration-Lifetime Constrained Power Grid Optimization Considering Multi-Segment Interconnect Wires. In *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*, ASP-DAC '18, pages 399–404, January 2018.

[73] Han Zhou, Zeyu Sun, Sheriff Sadiqbatcha, Naehyuck Chang, and Sheldon X.-D. Tan. EM-Aware and Lifetime-Constrained Optimization for Multisegment Power Grid Networks. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 27(4):940–953, April 2019.

[74] Han Zhou, Shuyuan Yu, Zeyu Sun, and Sheldon X.-D. Tan. Reliable Power Grid Network Design Framework Considering EM Immortalities for Multi-Segment Wires. In *Proceedings of the 25th Asia and South Pacific Design Automation Conference*, ASP-DAC '20, pages 74–79, January 2020.

[75] Qing K. Zhu. *Power Distribution Network Design for VLSI*. Wiley, Hoboken, NJ, 2004.