Electromigration Test Chip Design and Experimental Methodologies for Realistic Back-End-of-Line Failure Characterization and Physics-Based EM Aging Model Calibration

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA BY

Yong Hyeon Yi

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Chris H. Kim, Advisor

December 2023

Acknowledgments

First, I would like to thank my advisor, Professor Chris H. Kim. My Ph.D. life was full of valuable learning opportunities under his teaching and guidance. His suggestion of coming to the University of Minnesota for the degree was the best turnaround in my life. I appreciate his passionate commitment to fostering research excellence. The attitudes I learned from his mentorship altered my outlook and helped me successfully finish my degree.

I would also like to express my gratitude to my labmates. I will never forget the discussions and chats with our fellow researchers at the little Keller 4-147C office desk by the window. I cannot imagine finishing the degree without their help and support.

Finally, I am deeply grateful to my family for their support and encouragement. My parent's advice and love gave me the strength to continue the exciting but challenging long journey. I am truly fortunate to have such a wise, supportive family.

Dedication

To my loving family

Abstract

The reliability of BEOL (back-end-of-line) in a VLSI (very large-scale integration) system is an important concern both from a technology process and design perspective. In recent technology nodes, due to the scaling down of transistor feature sizes and metal routings, high current density and Joule heating that accelerates the interconnect aging are becoming serious reliability issues. Electromigration (EM), which is a focus of this thesis, is one major BEOL reliability concern. EM-induced material migration inside metal interconnects can create voids (local material vacancies) inside metal lines or vias that can increase local resistance or even open circuits in extreme cases. Due to this reason, semiconductor chip designers should understand the impact of EM on the power delivery networks and circuit interconnects to avoid or minimize aging and guarantee a chip's lifetime. Also, to build EM EDA (electrical design automation) tools for such EM lifetime verification, real silicon measurement data and statistics are required to validate and calibrate the models and algorithms of the software. Therefore, characterizing EM in different interconnect structures and experiment conditions is necessary, especially with the development of new technology nodes.

However, characterizing EM in a realistic structure is not a trivial task. For example, to characterize a power grid with multiple metal layers and various interconnect structures inside, monitoring the detailed EM behaviors in each net requires novel measurement methodologies. One solution for efficient measurement is building a test chip that contains EM test structures and measurement circuitries on the same die. Extensive data collection is available if the measurement circuitry is closely located to

the DUT (design under test) so that it can monitor the internal behavior of the complicated power nets. Also, by adding on-chip heaters and temperature sensors that help configure the EM experiment conditions to the same chip, the measurement efficiency is significantly improved. Given this context, in this thesis, EM test chip design and experiment methodologies employing on-chip heaters and measurement circuits are proposed. The focus of this work is an EM behavior inside a power grid or power grid-like structures that have complicated mesh structures with many redundancies and current paths.

The first part explains the cause of EM-induced aging and several models that can explain the interconnect failure behaviors. Also, based on the understanding of the EM mechanism, the physics-based EM simulation models that are calibrated by our silicon data will be introduced. The second part will propose the test chip experiment results for analyzing the impact of a temperature gradient on the power grid. This part will show not only the high temperature but also the 10s°C temperature gradient can change the grid lifetime significantly. The next part is an EM test chip experiment on realistic power grid structures. Four different specially designed realistic grids are tested under different currents and temperatures to compare different EM lifetime behaviors. The results are compared with the EM-induced stress evolution and voiding simulator to validate our silicon data. The last part focuses on the detailed advanced on-chip heater control techniques. The EM characterization in the proposed test chips cannot be done without the on-chip heaters for temperature acceleration and advanced control flow for accurate

data collection. To this end, the last part introduces the on-chip heater design/control details and possible difficulties during the EM test chip experiments.

Table of Contents

Chapter 1 Introduction 1
1.1 Impact of Temperature Gradient on Electromigration Lifetime 2
1.2 Electromigration Experiments From Realistic Power Grid Structures 3
1.3 On-Chip Heater Control Methodology 4
Chapter 2 Background 5
2.1 Electromigration-Induced Aging Mechanisms 5
2.2 Blech's Immortality Criterion 9
2.3 EM Time to Failure Model 10
2.4 Geometry Dependent EM Lifetime 13
2.5 Physics-Based EM Simulation 16
2.6 Temperature Gradient in Physics-Based Model 19
Chapter 3 Impact of Temperature Gradient in Electromigration Lifetime 23
3.1 Introduction 23
3.2 Test Chip Description 26
3.3 Experimental Methodology 28
3.3.1 Temperature Control 28
3.3.2 EM Failure Measurement 32
3.4 EM Failure Results and Discussions 33
3.4.1 Time to Failure Statistics 33
3.4.2 Failure Locations 35
3.4.3 Detailed Analysis and Comparison with EM Stress Simulator 37
Summary 39
Chapter 4 Electromigration Experiment From Realistic Power Grid Structure 40
4.1 Introduction 40
4.2 Test Chip Description 42
4.3 Experimental Methodology 44
4.4 Power Grid Failure Data Analysis Methodologies and Lifetime Prediction 46
4.4.1 Current vs. Voltage Stress 47

4.4.2 Temperature and Current Dependent Lifetime Behavior 50
4.5 Geometry-Dependent Failure Analysis and Model Calibration 53
Summary 59
Chapter 5 Advanced On-Chip Heater Control Methodology 60
5.1 Introduction 60
5.2 On-Chip Heater Design and Control Overview 60
5.3 Heater Temperature Control vs. DUT Temperature Control 64
5.4 Dual Threaded Control Flow 66
5.5 Multiple Heater Control and Multi-Variable Control Issue 68
5.6 Metal Heater Lifetime and Reliability Concerns 70
Summary 71
Chapter 6 Summary 72
Bibliography 75

List of Figures

Fig. 2.1. Electromigration-induced damage examples in a metal wire. Hillocks from material accumulation (left) and voids from depletion (right) [19] 5
Fig. 2.2. Cross section of a Cu wire indicating the back-stress and the electron wind force [20] 6
Fig. 2.3. Evolution of a stress distribution. (Top) the mechanical backstress starts to increase over time and reaches a steady state where electron wind induced atomic flux balances with the stress induced flux. (Bottom) If the tensile stress exceeds critical stress before reaching the equilibrium, void nucleation occurs [21] 7
Fig. 2.4. SEM (scanning electron microscope) images of (left) downstream electron and (right) upstream electron EM voids [22] 8
Fig. 2.5. An illustration of electromigration and stress migration until they reach the equilibrium. The equilibrium without reaching the critical stress (σ_{crit}) allows zero net atomic flux inside the wire, which allows immortality. The Maximum wire length L with a given current density j defines the Blech immortality condition, $(jL)_{Blech}$ [24] 10
Fig. 2.6. (Top left) Different mass diffusion paths. Bulk diffusion is the slowest since its activation energy (E_a) is the highest [25]. (Top right) The triple points at the crystal boundaries are usually the regions with material accumulations or depletions. (Bottom) Comparison of the crystal structures. The polycrystalline structures have many grain boundaries, so the grain boundary diffusion dominates EM. On the other hand, the bamboo structures with no continuous grain boundary paths are dominated by bulk diffusion (or lattice diffusion) [26] 12
Fig. 2.7. Interconnect design comparison. The left grid with wider metal lines and more via redundancy has a longer EM lifetime due to lower current density 13
Fig. 2.8. Interconnect lifetime comparison. (Left) Without the reservoir, void nucleation can undercut the via connection which can lead to an open circuit. (Right) By adding the material source at the bottom metal layer, it takes longer to completely block the current path [24] 14

Fig. 2.9. The role of source and sink in the reservoirs. The source-type reservoir works as additional material available and enables the development of voids without complete interruption of the current flow. The sink-type reservoir reduces the compressive mechanical stress by providing additional room for material accumulation [25] 14
Fig. 2.10. Simplified PDN example. An on-chip power grid consists of a tremendous amount of interconnects with complicated current distributions. Different types of EM failure scenarios coexist in such a large mesh structure 15
Fig. 2.11. Comparison of conventional and physics-based EM lifetime simulation on a metal mesh structure. 18
Fig. 2.12. EM silicon data from 9x9 metal mesh for validating the physics-based simulation model. (Left) First ten failures of a test chip. (Right) PDF of the first void nucleation time from the test chips [2] 19
Fig. 2.13. EM stress evolution trend comparison under different temperature gradient conditions. (a) The maximum stress location is the cathode side under uniform temperature conditions. (b-c) The maximum stress location gradually shifts right with the change of the maximum temperature gradient region [3]..... 20
Fig. 2.14. The grid lifetime statistics comparison of silicon data and simulations. (a) Uniform 350°C (623K) temperature condition. (b) With temperature gradient. (c) Temperauture gradient profile in (b) [3]..... 21
Fig. 3.1. Thermal gradient profile of a 7nm System on Chip (SoC)..... 23
Fig. 3.2. Tensile stress grows faster when a temperature gradient is present in a metal wire, due to the presence of larger atomic flux divergence 24
Fig. 3.3. EM test vehicle for studying temperature gradient effects within a chip which includes a 9x9 metal grid DUT, three poly heaters, and a 162:1 multiplexer for tapping out local voltages. 25
Fig. 3.4. (Left) Measured voltage drop of a fresh chip. The size of the arrows and dots denote voltage drops in the wires and vias. (Right) Each cross section has top and bottom tapping nodes. 26
Fig. 3.5. Test board and die photo 27

Fig. 3.6. (Top) TCR of three on-chip heaters measured from the same chip. (Bottom) TCR of all 38 tested DUTs used for extracting the average DUT temperature. 28
Fig. 3.7. (Top) Histogram of DUT temperatures for the 38 chips measured in this work. (Bottom) Average heater temperatures for each test mode were measured from 38 test chips. 29
Fig. 3.8. Experiment flow. 30
Fig. 3.9. (Upper left) Heater current control for the three heater mode. (Lower left) Heater temperatures for the three heater mode. (Right) Heater temperatures for the different modes 31
Fig. 3.10. EM data. (Upper) 192 voltage traces and (lower) DUT resistance versus stress time. 32
Fig. 3.11. Time to failure distributions for first, second, third, and fourth failures of each test condition. The two heater mode in green shows a bimodal behavior owing to the temperature gradient. For subsequent failures, the distributions revert to the standard lognormal shapes. 33
Fig. 3.12. (Upper left) Time to failure distribution of PCB #1, 2 heater mode highlighted in green. First EM void location of each of the 11 chips are shown in the order at which they occurred (e.g., first chip to fail denoted as '1', etc.). 34
Fig. 3.13. Location and occurrences of the first EM failures for each test condition. Early failures can be seen in the center figure along the bottom horizontal wire where the temperature gradient is the highest. 35
Fig. 3.14. On-chip heater and power grid layout and dimensions, drawn to scale. 36
Fig. 3.15. Simulated temperature map overlaid with the test structure showing the impact of temperature and temperature gradient on EM location. 36
Fig. 3.16. Comparison between three and two heater modes in terms of temperature gradient (from FEM tool), tensile stress (from EM tool), and via/wire voltage drops (measured). Tensile stress in node 62 increases beyond the critical level in the 2 heater mode, causing an EM failure despite the lower voltage and lower temperature. 37
Fig. 4.1. 28nm EM test chip overview. DUT includes a power grid (M2-M4) and quasi-cells with gate equivalent poly resistances. The heating area is covered by two metal

heaters (M5) and a temperature sensor (M2). The voltage measurement circuit includes two scan-chains and two 512:1 analog muxes. 41
Fig. 4.2. Breakdown of voltage tapping nodes and tapping locations in quasi-cells, power rings, and power rails. The tapping nodes allow measuring voltage delivery from rings and rails to each cell. 42
Fig. 4.3. Designed power grids' structural difference details. (a) Cell via counts. Min vs. max cell via (b) M2, M3 rail widths and M2-M3 vias. (c) M3, M4 rail widths and M3-M4 vias. (d) Number of M3 and M4 rails (rail density) 43
Fig. 4.4. Comparison table of four different DUTs. DUT2 is the baseline design with maximum via and rail strengths. 43
Fig. 4.5. Experiment flow with two concurrent controls. Thread 1 maintains consistent DUT temperature. Thread 2 runs only if the DUT temperature is fully stable. 44
Fig. 4.6. (Top left) Temperature control logs and heater current data. (Bottom left) Measured heater currents. (Right) Histogram of heater temperature under 350˚C sensor temperature. 45
Fig. 4.7. Temperature coefficient of resistance (TCR) of the heaters and the temperature sensor 46
Fig. 4.8. (a) Measured fresh chip DUT2 cell VDD & VSS voltage map. (b) Time zero cell VDD-VSS voltage map 47
Fig. 4.9. (a) Power grid resistance shifting profiles are plotted for TTF measurement. The TTF criterion is a 5% resistance shift. (b-c) Comparison of current and voltage stress. The ring, rail, and cell voltages are averaged (top). The voltage differences between the rings, rails, and cells for capturing subtle EM beginning times and locations (bottom). 47
Fig. 4.10. Error map (voltage shift) of each cell VDD-VSS at (a) fresh state, (b) beginning of EM, and (c) when the power grid fails (TTF). (d) The power delivery is gradually aggravated after the EM stress. 48
Fig. 4.11. Current & voltage stress average TTF (left) and TTF histogram (right) 49
Fig. 4.12. Time zero grid resistances at 350°C. (Left) Average resistances. (Right) Histogram of the resistance 49

Fig. 4.13. Power grid resistances and failure time comparison under (a) four temperatures and (b) four currents, respectively. 50
Fig. 4.14. Measured average TTFs. (Left) temperature dependent TTF. (Right) current dependent TTF 51
Fig. 4.15. TTF histograms. (Left) temperature dependent TTF. (Right) current dependent TTF 51
Fig. 4.16. Average TTFs are extrapolated for grid lifetime estimation. Black's equation (bottom) is assumed for the TTF calculation just for the simplicity of the lifetime analysis. 52
Fig. 4.17. DUT2&DUT3 TTF. Average TTF (left). TTF histogram (right) 53
Fig. 4.18. (a) TTF comparison of DUT3 and DUT2 (top) and DUT3's cell voltage profile (bottom). (b) Rings, rails, and cell voltage traces of DUT3 54
Fig. 4.19. Error maps of DUT3 (narrower rails) 54
Fig. 4.20. Measured TTFs of different types of grids 55
Fig. 4.21. Measured voltage traces at 350˚C. (a-c) 100mA stress current. (d) 10mA stress current. 55
Fig. 4.22. EM void simulation results for M2 and M3 rails. 57
Fig. 4.23. Voiding locations in upstream/downstream EM scenarios. 58
Fig. 4.24. The number of voids predicted by an industry-grade EM void simulator validates the earlier EM in DUT1 VSS nets. 58
Fig. 5.1. (Left) On-chip heater with four-terminal Kelvin testing connection. (Right) Multiple heaters can be used to raise the temperature of a larger circuit area. [18] 61
Fig. 5.2. Heater temperature control loop. [18] 61
Fig. 5.3. Temperature chamber (left). Board in the chamber with the 22nm EM test chip (metal heaters included) in the socket (right) 62
Fig. 5.4. (a) Vertical distance between temperature sensor and metal heaters, (b) Histogram of the sensor temperatures of 14 chips when the heater temperature is regulated at 400°C. 28nm process technology 64
Fig. 5.5. Histogram of the sensor temperatures of 10 chips when the sensor temperature is regulated at 350°C 65

Fig. 5.6. 28nm EM test chip experiment flow with two concurrent controls (top). The heater temperature of with the current direction switches every 20 seconds (bottom) 67
Fig. 5.7. (a) Thermal interferences between heaters can cause thermal runaway issues and prevent stable temperature control. (b) Measurement results show the leftmost heater dominating the temperature control. This issue was resolved by designing a new board with better thermal conductance and lowering the ambient temperature. 69
Fig. 5.8. (a) Metal heater resistance's change over its lifetime in a 28nm technology test chip. (b) Impact of heater degradation on temperature measurement accuracy. (c) Heater lifetime experiment. 70

Chapter 1 Introduction

Electromigration (EM) is a significant BEOL reliability concern in VLSI systems. In recent technology nodes, a semiconductor chip has tremendous amounts of interconnect structures with metal lines and vias in an on-chip power distribution network (PDN) and circuit interconnects. However, since the interconnects inside the chip are exposed to electron winds from DC (e.g., power supply nets) or AC current (clock or signal nets), a migration of the metal ions could result in vacancy or accumulation of the conductor mass, which could potentially result in open circuits (e.g., void in a wire) or unintended shorts (e.g., hillock and whisker). In terms of thin metal wire structures confined in metal liners or dielectric caps, which are common in VLSI chips, the void growing scenario is a major concern since it locally increases the resistance of the interconnect. The aging of the interconnect can result in timing variation in the circuit (i.e., additional circuit delay) or excessive IR drop that can lead to a power delivery failure. Furthermore, due to the reduction of the transistors and the interconnect sizes, EM is becoming a more serious aging concern in cutting-edge technology nodes. Therefore, chip designers have to verify such EM damages in their circuit layout to guarantee the semiconductor chips' lifetime.

Due to these reasons, there have been many approaches characterizing EM to predict its aging effect. A widely used characterization method is to raise the temperature of a device under test (DUT) using oven experiment setups for baking metal interconnects on silicon and apply a test stress current to estimate the lifetime of the

interconnect. However, even though oven-based EM characterization is straightforward, such experiments require expensive instrument setups and are usually limited to simpler structures due to failure measurement difficulties. On the other hand, by utilizing EM monitoring circuitry with on-chip heaters with precise localized heating systems, relatively complicated test structures (e.g., mesh structures with many redundancies) with efficient data collection are available. In this context, we propose on-chip EM characterization methodologies for monitoring metal interconnect's EM aging effect.

In addition to the better testability of the test chip-based EM monitoring approach, the data collected from the realistic/practical metal interconnect structures can contribute to the electronic design automation (EDA) society. This is because the development of an EM lifetime verification EDA software requires not only accurate aging physical models and algorithms but also actual silicon data to validate and calibrate the functionality of the software. In that respect, this work also compares the EM failure measurement and the state-of-the-art EM simulator to show how they can provide valuable data to the EDA world [2][3]. The details of the EM failure mechanisms, models, and physics-based simulators will be introduced in Chapter 2.

1.1 Impact of Temperature Gradient on Electromigration Lifetime

Chapter 3 is the test chip experiment and the analysis of the impact of temperature gradient on EM lifetime. In a conventional VLSI chip building workflow, EM violation checks are done on the metal routings in layouts to guarantee the product's lifetime.

Since the EM lifetime is a function of current density [1] (the lifetime model called Black's equation will be explained in the next chapter), the current density has to be limited to a certain criterion so as not to have EM-induced local resistance shift from the internal void growth. In most cases, such sign-off verifications are done under the assumption of uniform temperature. However, in reality, the die temperature of a chip cannot be uniform; it has multiple hot spots, resulting in multiple temperature gradient regions. Furthermore, the lifetime of an interconnect is not only the function of current density but also the temperature gradient and the stress distribution [4]. This is because the atomic flux of the material in a metal line is a combination effect of electrons, temperature, and mechanical stress. Therefore, to accurately characterize the EM-induced failure in a realistic scenario, the impact of temperature and mechanical stress should be considered and experimented. In this work, to compare the EM behavior with the impact of the 10s°C temperature gradient, we built a test chip with multiple on-chip heaters and a metal mesh structure to explain the EM failure location and the lifetime behavior.

1.2 Electromigration Experiments From Realistic Power Grid Structures

Chapter 4 is about the EM experiment on a realistic power grid structure. EM in an on-chip power delivery network is one of the biggest interconnect reliability concerns due to the huge DC current flowing from VDD to the ground. Compared with the AC current scenario, which is dominant in the interconnects in CMOS logic circuits whose

lifetime is significantly longer than the DC lifetime due to the healing effect [5], the power rails and vias exposed to a unidirectional current require careful designs and lifetime estimations. If EM voids grow inside a power grid and result in resistance shifts or even local open circuits, the excessive IR drop can lead to a power delivery failure to each internal logic cell. To this end, this work presents multiple power delivery network structures and experiments of the EM-induced aging in each of them under different currents and temperatures to analyze the geometry-dependent, stress current-dependent, and DUT temperature-dependent EM lifetime trends.

1.3 On-Chip Heater Control Methodology

Finally, an advanced on-chip heater control methodology will be discussed in Chapter 5. The aforementioned EM test chip uses polysilicon or metal line-based heaters to increase the die temperature at a local target heating area. The functionality of the heater is critical since EM lifetime has an exponential dependency on the absolute temperature. Therefore, to collect reliable statistical data, maintaining a heater or DUT temperature as accurately as possible is necessary. Such temperature consistency of the on-chip heaters requires well-calibrated control techniques to ensure the desired test conditions. Thus, in this section, detailed heater operating methodologies for the EM test vehicle are introduced.

Chapter 2 Background

2.1 Electromigration-Induced Aging Mechanisms

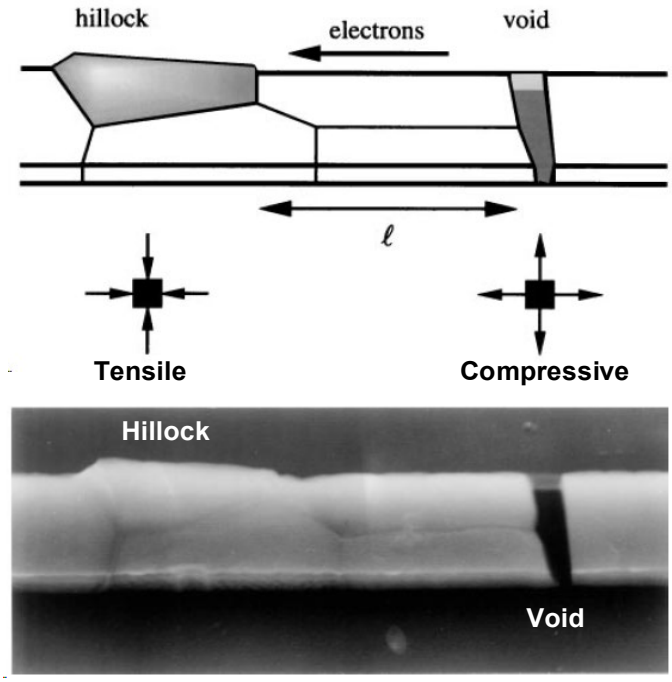


Fig. 2.1. Electromigration-induced damage examples in a metal wire. Hillocks from material accumulation (left) and voids from depletion (right) [19]

Electromigration is a BEOL aging effect resulting from the momentum transfer from electron wind to the metal ions (e.g., Cu, Al ions) in an interconnect that can induce material migration. Fig. 2.1 shows examples of EM-induced damage in metal wires. If Cu or Al atoms are deposited in the direction of the electron wind, the accumulation can generate an electrical short (Fig. 2.1 (left)). On the other hand, the depletion of the conductor atoms from the material migration can result in voids, which will increase the

wire resistance or even an open connection (Fig. 2.1 (right)). Since these aging concerns are no longer trivial in modern VLSI systems with high current density and high operating temperature, the EM failure has to be well characterized.

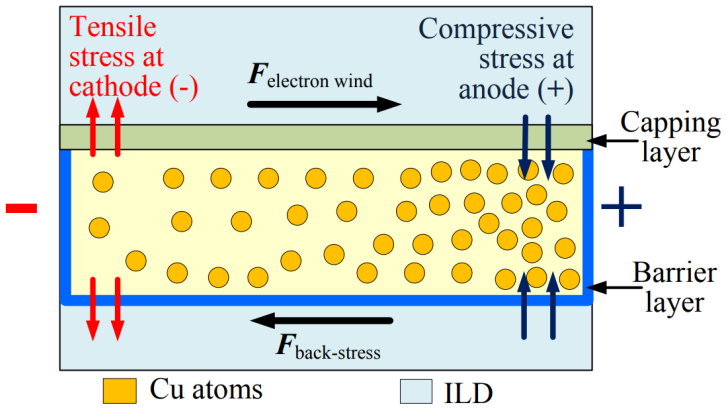


Fig. 2.2. Cross section of a Cu wire indicating the back-stress and the electron wind force [20]

The simplest way of characterizing EM and estimating metal interconnect lifetimes is experimentation on a single wire structure. For such characterization, a basic metal wire structure described in Fig. 2.2 is normally assumed. In this metal wire cross-section illustration, electron wind moves from the cathode (left) to the anode (right). The wire is encapsulated by the barrier layer (e.g., Ta/TaN, conductive liner) and the capping layer (e.g., SiN, NSiC, dielectric cap). Due to the confined structure, the Cu atom cannot move outside the boundaries. Since the Cu atoms move toward the electron wind direction (cathode to anode), the cathode side gradually gets depleted while the anode region experiences mass accumulation. At the same time, such gradual depletion of the cathode generates an increase of tensile stress, whereas the anode's accumulation results in compressive stress, forming new stress distribution with gradients. For example, the

maximum compressive stress at the anode pushes the Cu atoms to the cathode side,

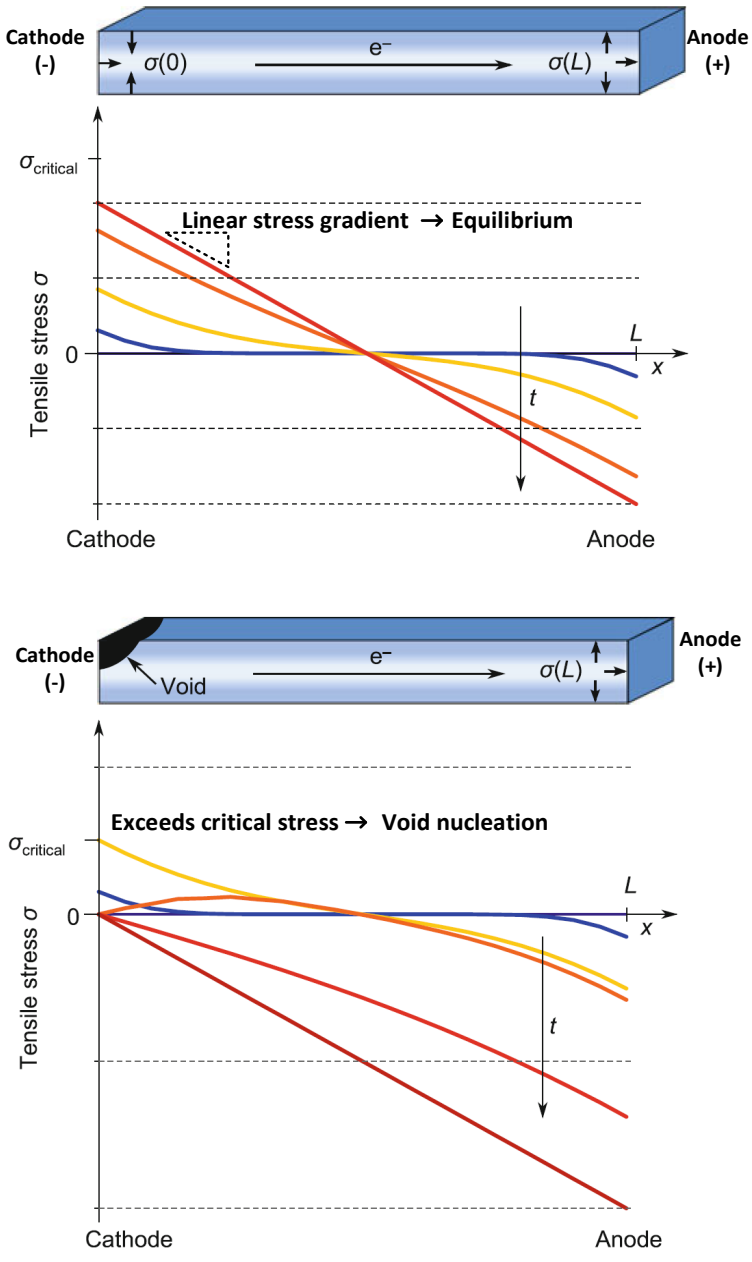


Fig. 2.3. Evolution of a stress distribution. (Top) the mechanical backstress starts to increase over time and reaches a steady state where electron wind induced atomic flux balances with the stress induced flux. (Bottom) If the tensile stress exceeds critical stress before reaching the equilibrium, void nucleation occurs [21]

where the maximum tensile stress is from relative vacancy. Thus, the material migration from the mechanical back-stress ($F_{back\text{-}stress}$) occurs in the opposite direction, starting to counterbalance the electron wind-induced atomic flux.

Fig. 2.3 (top) is the illustration of the stress distribution during EM. Since the electric field applied to the wire is fixed, the atomic flux from the electron wind is constant. However, the stress gradient starts to grow due to the atomic density change and eventually reaches a steady state with a constant stress gradient in the end. In this equilibrium state, the atomic flux from the electrons and the stress distribution balance each other, and the net atomic flux becomes zero along the wire. On the other hand, as shown in Fig. 2.3 (bottom) scenario, if tensile stress at one of the locations exceeds a critical stress (σ_{crit}) before reaching the equilibrium, void starts to form, which is called void nucleation. For this wire, since the tensile stress at the cathode side is at its

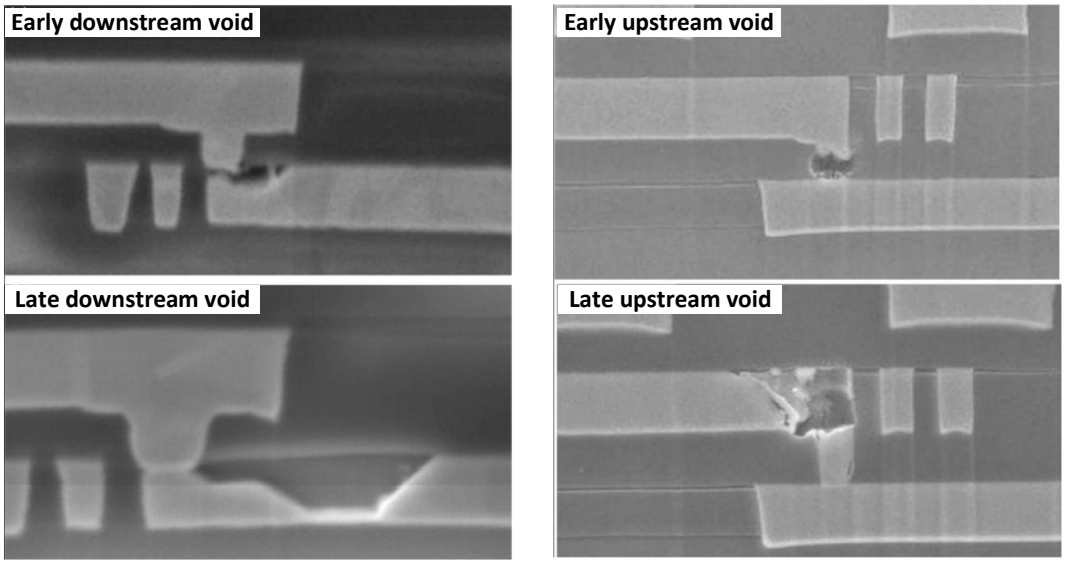


Fig. 2.4. SEM (scanning electron microscope) images of (left) downstream electron and (right) upstream electron EM voids [22]

maximum, a void occurs in the cathode region. Since the void is formed, the tensile stress disappears at the cathode.

If the explained void nucleation occurs nearby or inside a via interconnect that connects the upper metal and the lower metal layers, the void can result in an open circuit, as shown in Fig. 2.4. For example, if electrons flow from the upper to the lower layer, as shown in downstream void images (left), the void starts to grow under the via, which is a local cathode region. Similarly, the upstream void occurs in or above the via, as shown in Fig. 2.4 (right). Both permanent damage can result in significant power delivery or circuit interconnect failures. Note that power grids with unidirectional current have more aging concerns, which is a major focus of this thesis.

2.2 Blech's Immortality Criterion

As explained in the previous EM voiding mechanism, void nucleation happens if the EM and SM (stress migration) cannot balance and fail to become an equilibrium state before reaching the critical stress. In this context, a concept called Blech length [23] is widely used to define an immortal metal wire length. As illustrated in Fig. 2.5, with a given current density in a wire segment, the maximum wire length L that allows the zero net atomic flux (i.e., equilibrium) can be experimentally defined. The product of the current density and the maximum wire length for the wire immortality is called "Blech's criterion". Following is the wire segment's immortality condition defined by the product of current density (j) and the wire length (L).

$$(jL) < (jL)_{Blech} \tag{1}$$

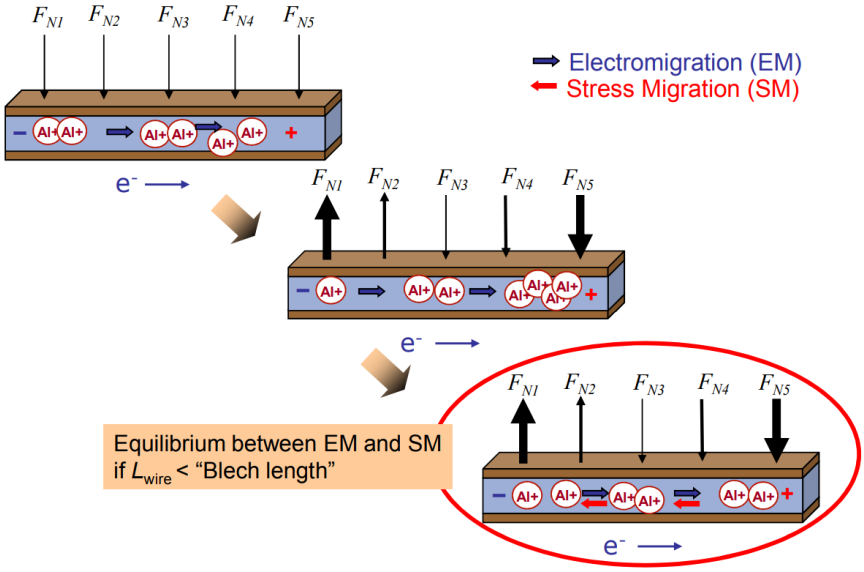


Fig. 2.5. An illustration of electromigration and stress migration until they reach the equilibrium. The equilibrium without reaching the critical stress (σ_{crit}) allows zero net atomic flux inside the wire, which allows immortality. The Maximum wire length L with a given current density *j* defines the Blech immortality condition, $(jL)_{Blech}$ [24]

2.3 EM Time to Failure Model

One of the most widely accepted EM failure models is Black's equation [1]. Black's equation is a widely used median-time-to-failure (MTTF) model that can explain the time-to-failure of a metal wire segment straightforwardly.

$$MTTF = \frac{A}{j^n} \cdot \exp\left(\frac{E_a}{k \cdot T}\right) \quad (2)$$

In the equation, j is the current density, k is the Boltzmann constant, T is absolute temperature, E_a is the activation energy, and A is a wire cross-section-dependent constant. The current density exponent n and the E_a are material-dependent parameters. One should note that the failure time of a metal line depends on the current density (j) and the

temperature (T). Due to this relationship, the EM experiment is normally done under high current density, and high temperature conditions to accelerate the aging to collect the data in a reasonable amount of time. Once the lifetime of the wire is measured at the high temperature and high current, the lifetime could be extrapolated to the temperature and current density levels of real chip operating conditions (i.e., lower temperature, lower current density). Note that the model only supports single metal line interconnects and has limitations on predicting the lifetime of realistic interconnects with many redundancies, such as power delivery networks (PDN).

The activation energy (E_a) parameter also helps in understanding the EM behavior. Intuitively, E_a is a resistance of the metal ions to diffuse. For example, if E_a is high, it is more difficult to induce material migration, which is translated into slower mass diffusion. The activation energy depends not only on the conductor material itself but also on the crystal structure and grain boundaries. Fig. 2.6 (top left) shows different types of grain boundaries. Among the three diffusion paths, the bulk diffusion, which is a diffusion inside the crystal lattice, has the largest activation energy. The diffusion path at the grain boundaries and the surface have lower E_a meaning mass diffusion along those regions is faster than the bulk of a crystal. Therefore, if a metal line has fine-grain crystals with many grain boundaries (polycrystalline), the grain boundary and surface diffusion dominate EM. The polycrystalline structure's dominance also comes from the effect of triple points. Since the material depletion or accumulation, which potentially leads to the generation of voids or hillocks, comes from the divergence of the diffusion, the triple point at the boundaries experiences more material density changes (Fig. 2.6 (top

right)). In a thin metal wire which is common in interconnects in a VLSI system, polycrystalline with many grain boundaries and triple points is a general structure (Fig. 2.6 (bottom)). However, if the wire is thin enough to form a bamboo or near bamboo structure, due to the lack of continuous grain boundary paths (i.e., the vertical bamboo structure blocks the diffusion), the bulk diffusion or lattice diffusion dominates the migration. In that scenario, the metal wire's EM lifetime is improved compared to the polycrystalline structure. Since there are multiple diffusion paths with different E_a values, the effective diffusion coefficients (E_a^{eff}) are often defined as combinations of bulk, surface, and grain boundary diffusion coefficients.

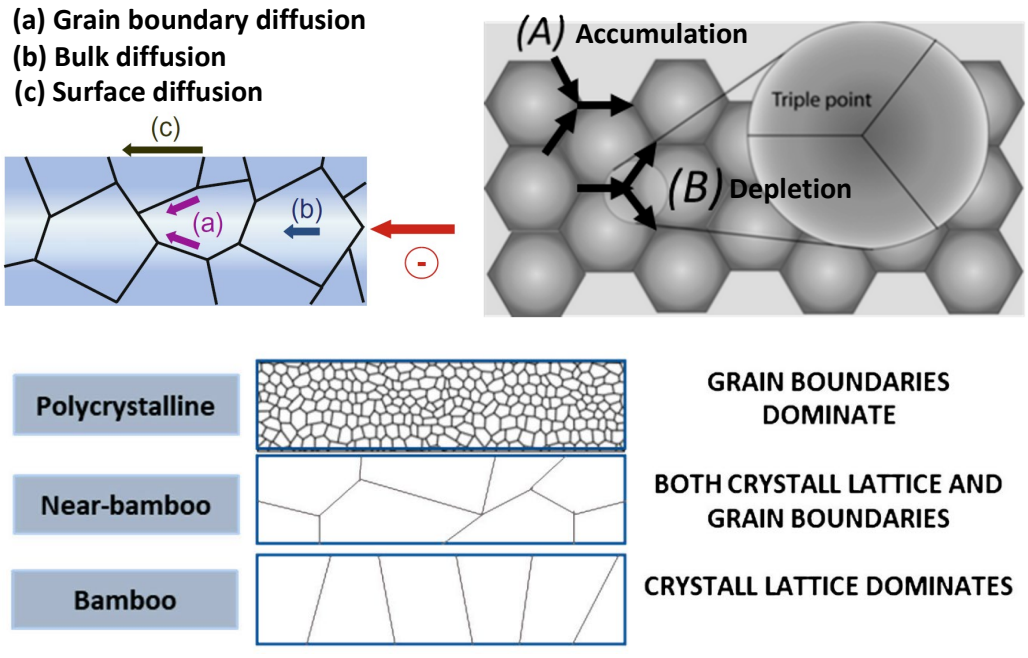


Fig. 2.6. (Top left) Different mass diffusion paths. Bulk diffusion is the slowest since its activation energy (E_a) is the highest [25]. (Top right) The triple points at the crystal boundaries are usually the regions with material accumulations or depletions. (Bottom) Comparison of the crystal structures. The polycrystalline structures have many grain boundaries, so the grain boundary diffusion dominates EM. On the other hand, the bamboo structures with no continuous grain boundary paths are dominated by bulk diffusion (or lattice diffusion) [26]

2.4 Geometry Dependent EM Lifetime

Since the major cause of the EM failure is a material migration, the failure time highly depends on the geometry of the interconnects. One simple way of mitigating the interconnect failure is minimizing the current density in the wires and vias. As explained in Black's equation (expression (2)), because the EM TTF is a function of current density, to mitigate EM failure, designers should decide the metal routing layouts in the direction of increasing the wire cross-section area and maximizing the usage of the vias. The easiest design decision is to increase the width of the metal line and use more via redundancy. For example, as shown in Fig. 2.7, if allowed, the left grid design is recommended for better interconnect lifetime.

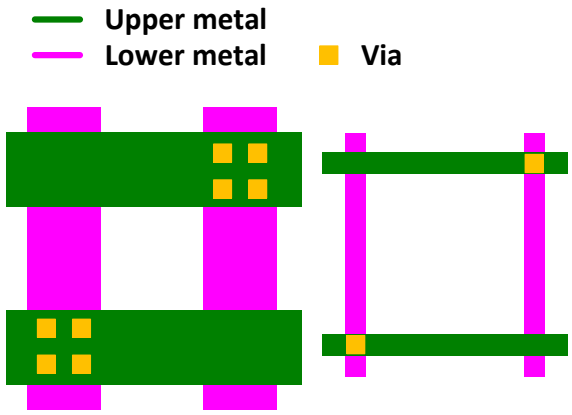


Fig. 2.7. Interconnect design comparison. The left grid with wider metal lines and more via redundancy has a longer EM lifetime due to lower current density

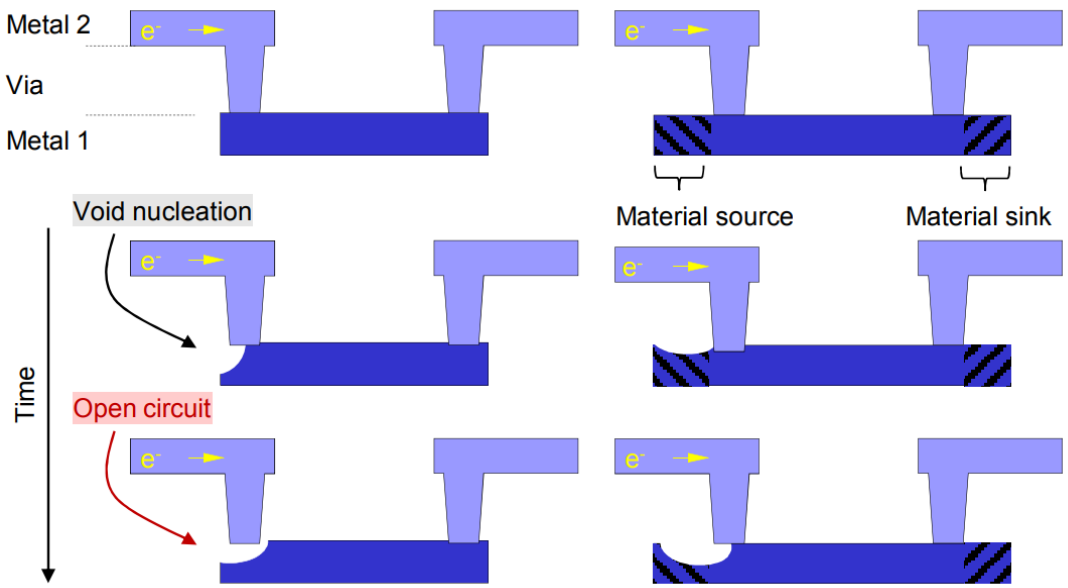


Fig. 2.8. Interconnect lifetime comparison. (Left) Without the reservoir, void nucleation can undercut the via connection which can lead to an open circuit. (Right) By adding the material source at the bottom metal layer, it takes longer to completely block the current path [24]

Another significant lifetime improvement can be made by utilizing a reservoir effect. Fig. 2.8 explains how the enlarged metal wire can change the voiding behavior. As shown in Fig. 2.8 (right), the additional metals near the via regions work as a material

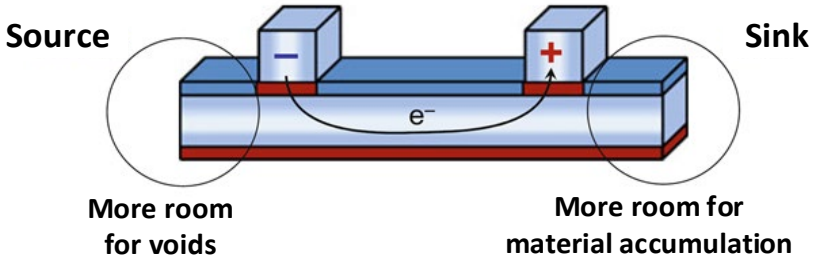


Fig. 2.9. The role of source and sink in the reservoirs. The source-type reservoir works as additional material available and enables the development of voids without complete interruption of the current flow. The sink-type reservoir reduces the compressive mechanical stress by providing additional room for material accumulation [25]

reservoir that needs to be drained before void nucleation starts right underneath the via. Due to the material source, the impact of voiding could be mitigated, and the time to reach the complete open circuit is significantly improved. The role of the source and sink reservoirs are described in Fig. 2.9. The source reservoir provides additional room for voiding and can prevent the early below via voiding that can completely block the current flow. On the other hand, the sink reservoir works as a room for material accumulation, which can release the compressive stress at the anode region. Note that the hillocks and whiskers from the material accumulation are less concern than voids since the critical stress for their creation is higher than that of voids.

Modern VLSI systems have complicated interconnect structures. Therefore, the introduced interconnect lifetime estimation for a wire segment cannot accurately predict the aging effect of the metal mesh with many redundancies and complex current

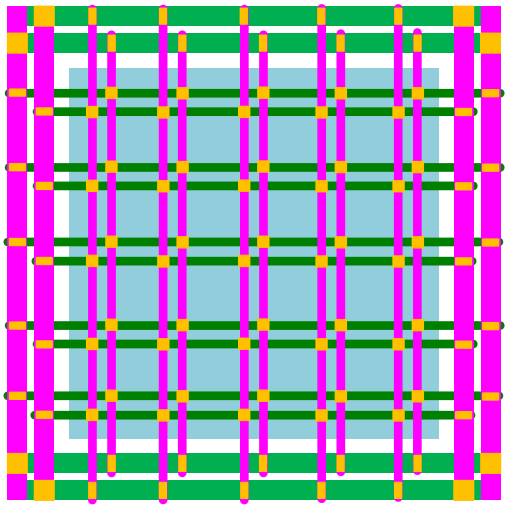


Fig. 2.10. Simplified PDN example. An on-chip power grid consists of a tremendous amount of interconnects with complicated current distributions. Different types of EM failure scenarios coexist in such a large mesh structure

distributions such as PDN (Fig. 2.10). To this end, more sophisticated models are required to estimate the interconnect lifetime holistically.

2.5 Physics-Based EM Simulation

Previous sections (sections 2.2 and 2.3) explained how the metal line's critical length (Blech length) and the MTTF (Black's equation) are characterized to estimate the lifetime of the wire. Based on that information, once a layout of the metal routings is carried out, the EM violation check could be done by calculating the amount of current (or current density) in each wire segment. With the wire geometry (width, length, thickness) and the calculated current inside, a simulator can estimate the lifetime of each interconnect segment. If the lifetime of the segment cannot fulfill the system's target lifetime (e.g., 10 years) designers should change the geometry of the routings to reduce the current density and resolve the EM violations. However, as introduced in section 2.4, EM failure highly depends on the geometry of the interconnect branches since the redundancy of the wire nets can significantly improve the lifetime (e.g., reservoir effect). Therefore, the conventional EM validation can underestimate the lifetime of the metal wire nets and possibly encourage unnecessary usage of routing resources, which is translated to an increase in the overall area required.

A physics-based EM simulation methodology can capture the lifetime of the interconnect nets more precisely by focusing on material migration and the stress evolution along the nets. One of the most frequently used EM-induced stress evolution models is Korhonen's equation [27].

$$\frac{\partial \sigma}{\partial t} = \frac{\partial}{\partial x}\left[\frac{D_a B \Omega}{kT}\left(\frac{\partial \sigma}{\partial x} - \frac{z^* e \varrho}{\Omega} j\right)\right] \tag{3}$$

Here, σ is the hydrostatic stress, t is the time, D_a is the atomic diffusivity, B is the applicable modulus, Ω is atomic volume, k is Boltzmann's constant, T is absolute temperature, z^* is the effective charge number, e is the electron charge, ϱ is resistivity, and j is the current density. The 1-dimensional equation explains the stress evolution at each location over time in a confined thin metal wire. An intuitive way to understand the model is to understand the flux components inside the equation (3). There are three major material migration components, according to [4].

$$\overrightarrow{J_E} = \frac{c}{kT} \cdot D_a \cdot z^* e \varrho j \tag{4}$$

$$\overrightarrow{J_T} = -\frac{cQ}{kT^2} \cdot D_a \cdot \nabla T \tag{5}$$

$$\overrightarrow{J_S} = -\frac{c\Omega}{kT} \cdot D_a \cdot \nabla \sigma \tag{6}$$

$$D_a = D_0 \cdot exp(-\frac{E_a}{k \cdot T}) \tag{7}$$

From expression (4-6), the c is the concentration of atoms, Q is the transported heat. In expression (7), D_0 is the diffusion coefficient at room temperature. Note that the diffusion coefficient D_a is a function of temperature. The atomic flux $\overrightarrow{J_E}$, $\overrightarrow{J_T}$ and $\overrightarrow{J_S}$ are from electromigration, thermomigration, and stress migration, respectively. The EM-induced atomic flux $\overrightarrow{J_E}$ is proportional to the current density, whereas the $\overrightarrow{J_T}$ is a function of temperature gradient and $\overrightarrow{J_S}$ is proportional to the stress gradient. Intuitively, if the net atomic flux $\overrightarrow{J_{net}} = \overrightarrow{J_E} + \overrightarrow{J_T} + \overrightarrow{J_S}$ is constant at every location, the material concentration

doesn't change since the influx and the outflux are identical. However, the material concentration occurs with the divergence of a net flux. Such material concentration change results in the stress evolution, as shown in Fig. 2.3 in section 2.1. In this context, the Korhonen's equation could be explained as the stress evolution resulting from the sum of $\overrightarrow{J_E} + \overrightarrow{J_S}$ with the assumption of uniform temperature (i.e., zero temperature gradient). Note that the term $\frac{\partial \sigma}{\partial x}$ and $\frac{z^* e \varrho}{\Omega} j$ in (3) are proportional to the $\overrightarrow{J_S}$ and $\overrightarrow{J_E}$, respectively.

The Korhonen's equation could be applied not only to a single wire segment but also to complicated wire nets. V. Sukharev et al. used this efficient 1D stress evolution analysis to simulate the EM failure behavior of a metal mesh structure [2]. By assuming continuous atomic flux and hydrostatic stress in adjacent branches, the time and location of void nucleations and the growths of the voids could be accurately simulated (Fig. 2.11).

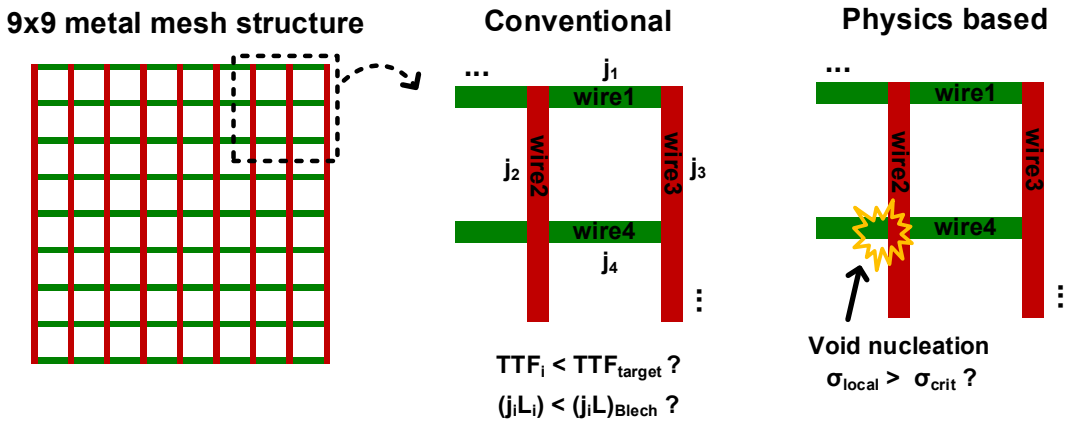


Fig. 2.11. Comparison of conventional and physics-based EM lifetime simulation on a metal mesh structure.

One of the challenges of validating the physics-based simulation is a lack of silicon data. The best way to calibrate the model in the EM simulator is to collect the EM failure information on a test chip using the same structure in the simulation. For example, in a simulation model in [2], effective activation energy values (E_a^{eff}) and critical stresses (σ_{crit}) are randomly decided at each branch, assuming normal distribution and Weibull distribution, respectively. The detailed parameters of the distributions could be decided by the MTTF information from the silicon experiments (Fig. 2.12 (right)) to match the aging behavior with the best fit. Also, the location of the void nucleations and the void growth/movement can be validated by the real EM failure locations and orders in a DUT (Fig. 2.12 (left)).

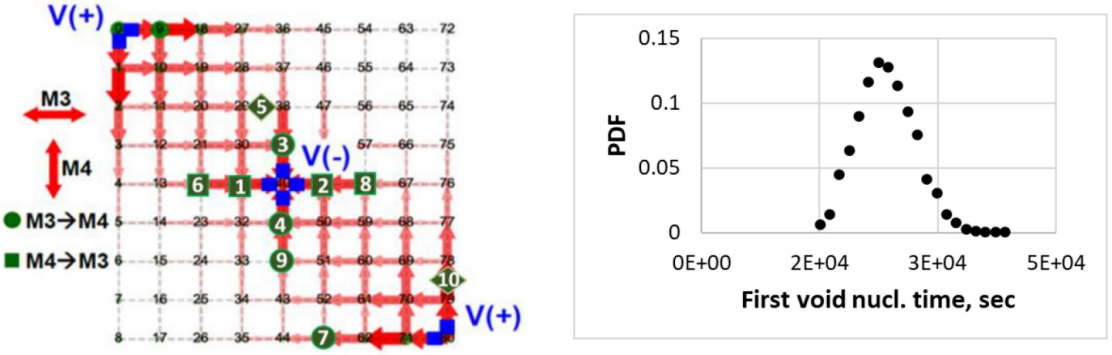


Fig. 2.12. EM silicon data from 9x9 metal mesh for validating the physics-based simulation model. (Left) First ten failures of a test chip. (Right) PDF of the first void nucleation time from the test chips [2]

2.6 Temperature Gradient in Physics-Based Model

Korhonen's equation (3) introduced in the previous section doesn't focus on the effect of temperature gradient. However, as shown in thermal atomic flux expression (5), the gradient of the temperature can contribute as an additional atomic flux component.

Also, as explained in expression (7), the diffusion coefficient is an exponential function of absolute temperature *T*. This is because the material diffusion becomes faster with higher temperature. Therefore, the existence of temperature gradient along interconnects cannot be ignored since it can create additional flux divergence, which can cut down the

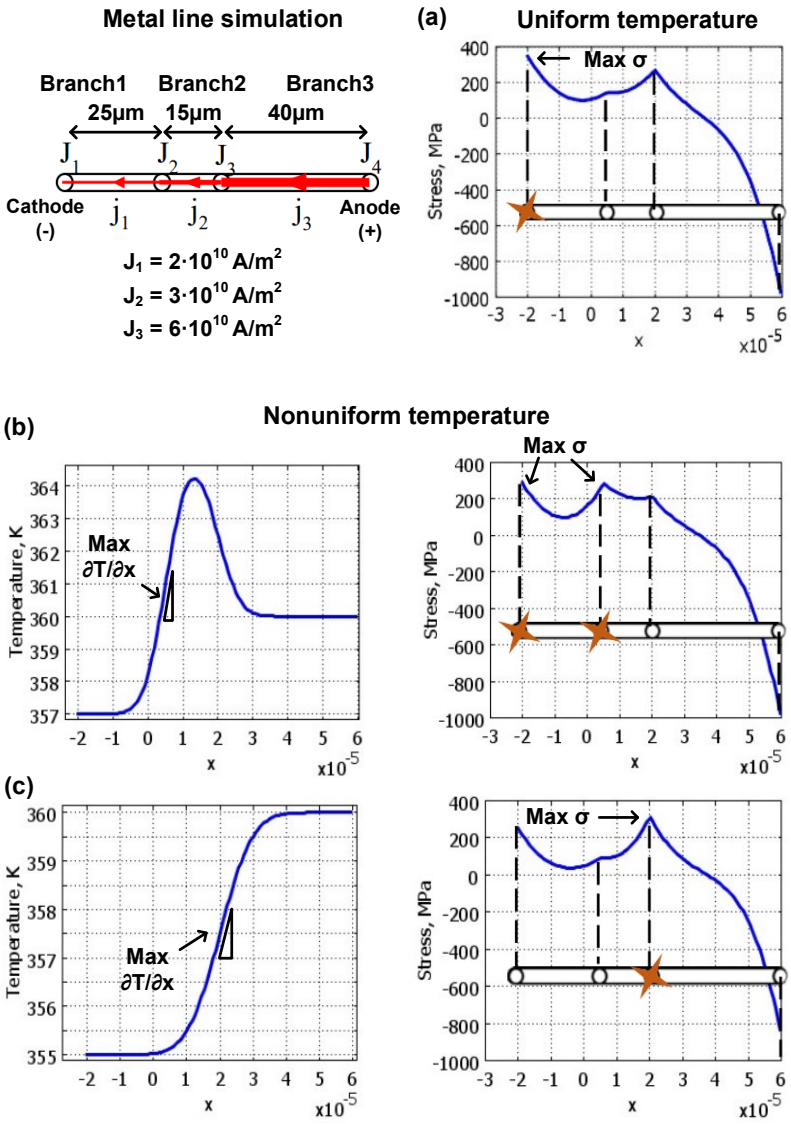


Fig. 2.13. EM stress evolution trend comparison under different temperature gradient conditions. (a) The maximum stress location is the cathode side under uniform temperature conditions. (b-c) The maximum stress location gradually shifts right with the change of the maximum temperature gradient region [3]

time to reach the critical stress, potentially leading to earlier EM failures. The impact of temperature gradient on the EM failure behavior is explained in Fig. 2.13. In this simulation scenario, at uniform temperature condition (Fig. 2.13 (a)), the metal wire's cathode side has maximum tensile stress due to the depletion of the mass. However, in Fig. 2.13 (b-c), the location of the maximum stress shifts with the maximum temperature gradient and changes the first void nucleation region. For instance, in Fig. 2.13 (b), the atomic diffusivity at branch2 is higher than branch1 (i.e., faster migration in branch2), and the additional outflux toward the right side increases the stress between branch1 and branch2. Similarly, in Fig. 2.13 (c), the diffusivity in branch3 is higher than in branch2, and the maximum tensile stress location is between branch2 and branch3.

In this context, A. Kteyan et al. [3] proposed the EM simulation model, including the effect of temperature gradients using the same DUT structures presented in [2] (Fig.

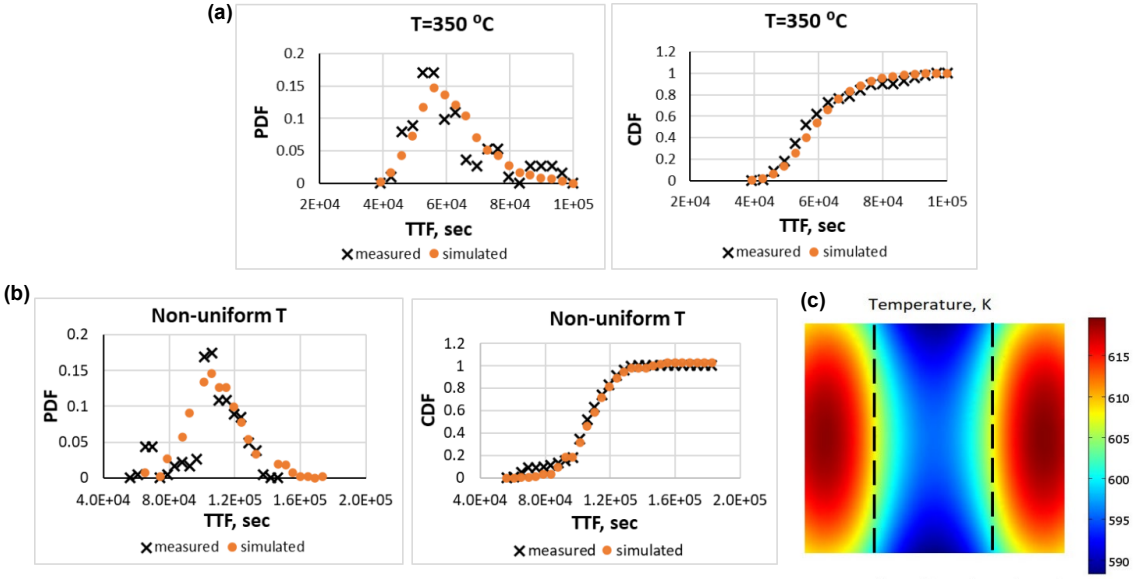


Fig. 2.14. The grid lifetime statistics comparison of silicon data and simulations. (a) Uniform 350°C (623K) temperature condition. (b) With temperature gradient. (c) Temperauture gradient profile in (b) [3]

2.12) with different temperature gradient conditions. The model could be improved if it is calibrated to capture the grid's lifetime distribution both in uniform temperature and temperature gradient conditions using measured silicon lifetime data (Fig. 2.14).

Chapter 3 Impact of Temperature Gradient in Electromigration Lifetime

3.1 Introduction

EM in power grids is a critical reliability concern due to the short DC stress lifetime and excessive IR drops caused by EM voids which may lead to power delivery failures. While state-of-the-art EDA tools are currently being developed to more accurately predict the EM lifetime of power grids, only limited silicon data is available to verify the accuracy of the power grid EM models and simulation tools. As part of the extensive EM silicon data collection, previous works have shown on-chip heater based AC [6-7] and DC [8] lifetime characterization results, as well as signal behaviors in damaged circuit interconnects [9-10]. However, such EM test structures are limited to metal wire segments rather than power-grid-like mesh structures with multiple redundant current paths. Several other works presented voltage evolution at various locations in a power grid, which helped reveal the true nature of EM voids in a complex mesh structure

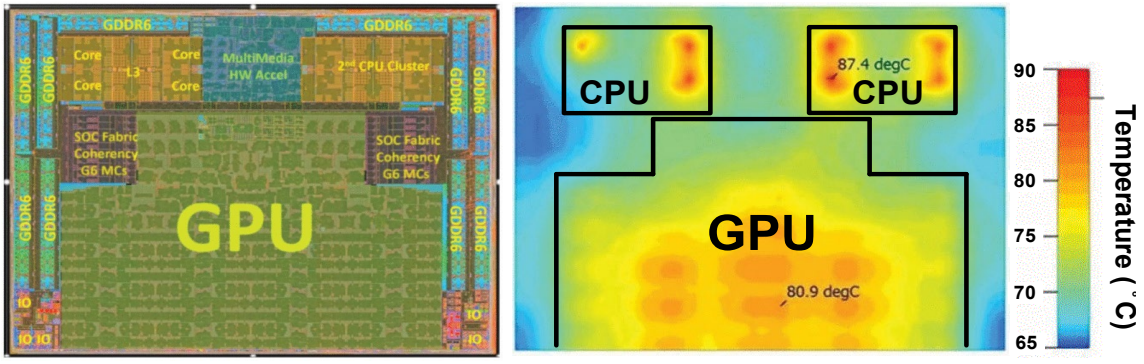


Fig. 3.1. Thermal gradient profile of a 7nm System on Chip (SoC).

with redundant/alternative current paths [11-13]. These efforts have helped build confidence in the EM models as well as our collective understanding of EM and will facilitate the adoption of new EM tools by the semiconductor industry [2]. However, a critical shortcoming of these works is that most EM data were collected under a spatially uniform (or close to uniform) stress temperature, which is not representative of a real power grid. Instead, a real power grid is exposed to thermal hot spots and local Joule heating effects, as shown in the thermal gradient profile in Fig. 3.1 [14].

Such non-uniform temperature accelerates a void formation in metal wires due to the faster tensile stress growth resulting from larger atomic flux divergence (Fig. 3.2).

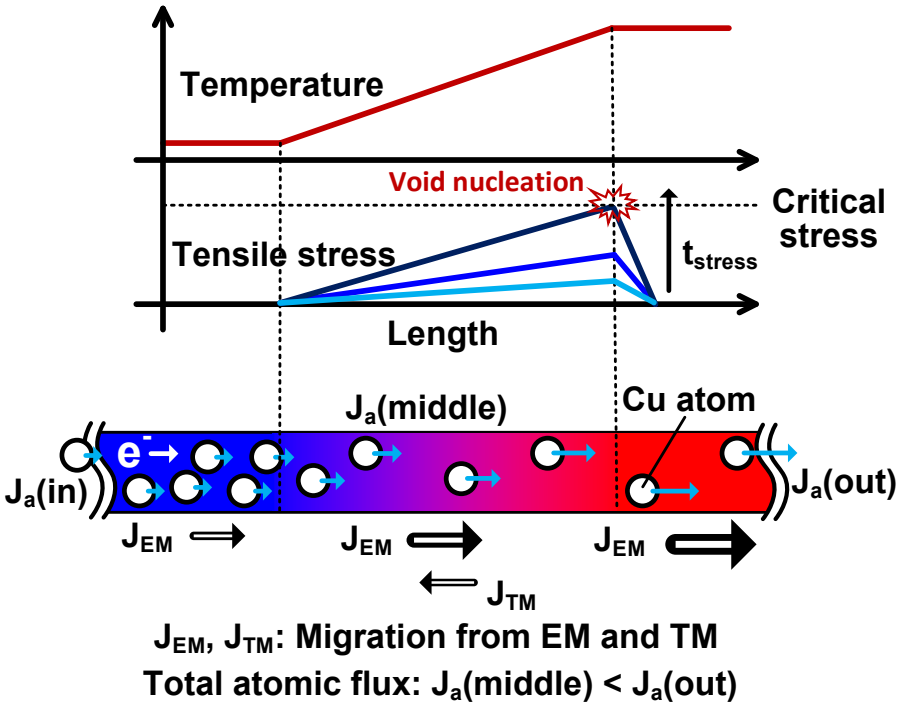


Fig. 3.2. Tensile stress grows faster when a temperature gradient is present in a metal wire, due to the presence of larger atomic flux divergence

In this work, we carried out the first-of-its-kind experiments to study the impact of local temperature gradient on the power grid EM lifetime. The test vehicle comprises a custom-designed 9x9 M3-M4 power grid mesh, three integrated heaters, and a voltage tapping scheme capable of measuring the voltage at each intersection of the 2D mesh. The temperature gradient, along with the stress temperature needed for accelerating EM, was applied by individually controlling the power of the three integrated heaters. The average temperatures of the three heaters and the power grid were monitored using the temperature coefficient of the resistance (TCR) method. The time to failure (TTF) distributions were collected under three test modes to understand the impact of temperature gradient on the failure location, failure order, and failure type. To our knowledge, this work presents the first experimental evidence of temperature gradient-induced EM failures in a power grid like test structure.

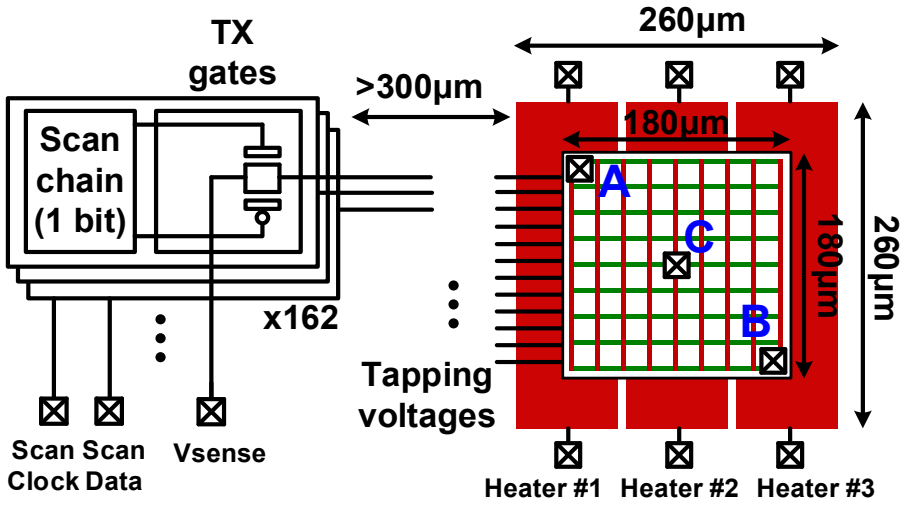


Fig. 3.3. EM test vehicle for studying temperature gradient effects within a chip which includes a 9x9 metal grid DUT, three poly heaters, and a 162:1 multiplexer for tapping out local voltages.

3.2 Test Chip Description

The basic overview of the EM monitoring test vehicle is shown in Fig. 3.3. The DUT is a metal mesh structure composed of horizontal M3 and vertical M4 layers. The width of the wires is 0.1μm (minimum), where the length of each branch segment is 20μm (Fig. 3.4 (right)). The cross-sections of the metals are connected by single M3 and M4 vias. Each cross-section has two voltage tapping nodes above and below the via that enable direct voltage measurement. The 9x9 grid has 162 tapping points in total, and the internal nodes are connected to the IO device-based voltage scanning circuit. The scanning circuit is far away from the heating area (> 300μm) to minimize the leakage current that may degrade the measurement accuracy. The three on-chip heaters are located beneath the DUT to raise the die temperature beyond 350°C. The silicided poly heaters are separately controlled, enabling a configuration of different temperature gradient conditions. For example, if the three heaters have the same temperature, the

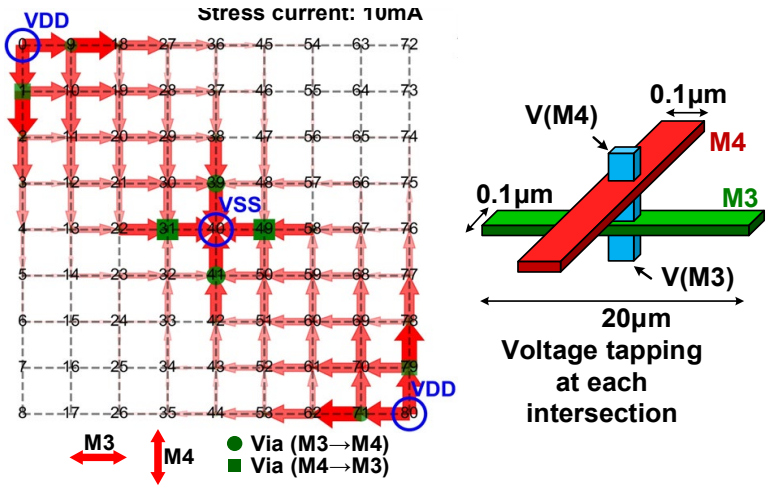


Fig. 3.4. (Left) Measured voltage drop of a fresh chip. The size of the arrows and dots denote voltage drops in the wires and vias. (Right) Each cross section has top and bottom tapping nodes.

lateral thermal distribution is almost uniform, whereas operating only the side heaters (heater #1, #3) forces the thermal gradient between the side and the middle area of the DUT.

Fig. 3.4 (left) is the voltage drop map of the fresh chip after a 10mA stress current is applied. The magnitude of the arrow on the segments and the dots on the vias denote the relative voltage drop throughout the mesh. This profile mimics a current distribution scenario of power grids in an IC block that contains complicated branches with numerous redundant paths. Since the temperature distribution throughout the die is highly affected by the heat sinking capability, PCBs (Fig. 3.5 (left)) from two separate vendors were ordered to measure any difference in the failure trend.

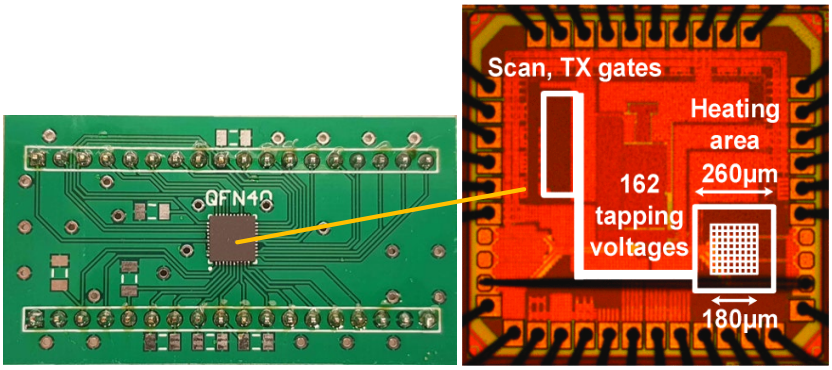


Fig. 3.5. Test board and die photo

3.3 Experimental Methodology

3.3.1 Temperature Control

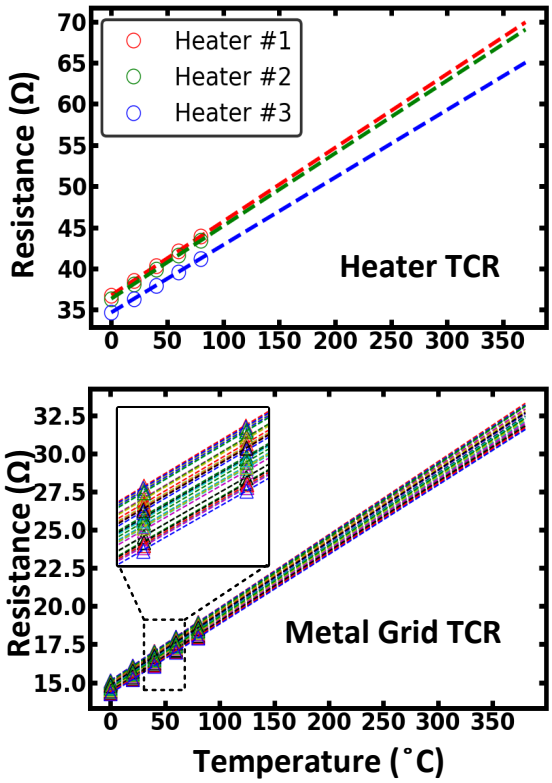
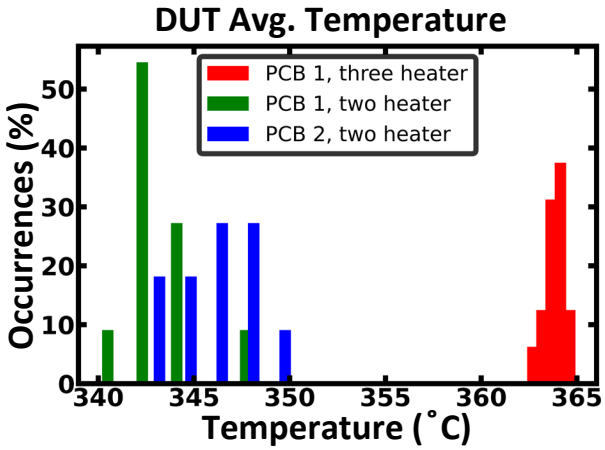


Fig. 3.6. (Top) TCR of three on-chip heaters measured from the same chip. (Bottom) TCR of all 38 tested DUTs used for extracting the average DUT temperature.

EM testing under various temperature gradient conditions requires accurate control and temperature tracking of the individual heaters. Due to this requirement, the TCR of the three heaters and the power grid DUT are characterized prior to the experiment. The TCRs of the heaters (top) and the DUTs (bottom) in Fig. 3.6 confirm an R^2 value of 0.9999 or higher. We measured the temperatures of DUTs and heaters based on the linear TCR extrapolation, which allows the local temperature to be estimated by measuring the resistance values.

Fig 3.7 shows the DUT and heater temperatures recorded before the stress for the three test modes. The temperature difference between the side heaters and the middle heater is 0°C, 30°C, and 20°C, respectively, for the three test modes. The number of chips tested for each mode is 16, 11, and 11, respectively. The two heater modes represent a realistic chip operating scenario where thermal distributions are nonuniform due to the local hotspots. Also, the histogram of the DUT temperature is presented in Fig. 3.7



Heater Temperature

Test mode	Heater #1	Heater #2	Heater #3
PCB vendor #1 Three heater (16 chips)	350.8˚C	350.7˚C	350.7˚C
PCB vendor #1 Two heater (11 chips)	350.8˚C	323.1˚C	350.7˚C
PCB vendor #2 Two heater (11 chips)	350.2˚C	332.3˚C	349.9˚C

Fig. 3.7. (Top) Histogram of DUT temperatures for the 38 chips measured in this work. (Bottom) Average heater temperatures for each test mode were measured from 38 test chips.

(bottom), which shows the average and variation of the DUT temperature for the three different modes.

Note that the measured temperatures of the DUTs (>360°C) are higher than that of the heaters (350°C) for the uniform temperature mode. We suspect this error comes from the nonlinearity of the heater TCRs at the high temperature region. Due to the combined resistivity of the p+ polysilicon (negative TCR) and the silicide (positive TCR) in the heaters, perfect linearity of the silicided poly heater cannot be guaranteed at extreme temperatures. Such deviation from the linear trend can underestimate the die temperature, which explains the higher DUT temperatures in the histogram.

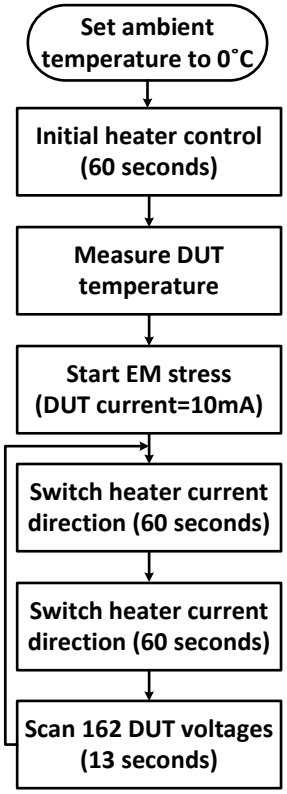


Fig. 3.8. Experiment flow.

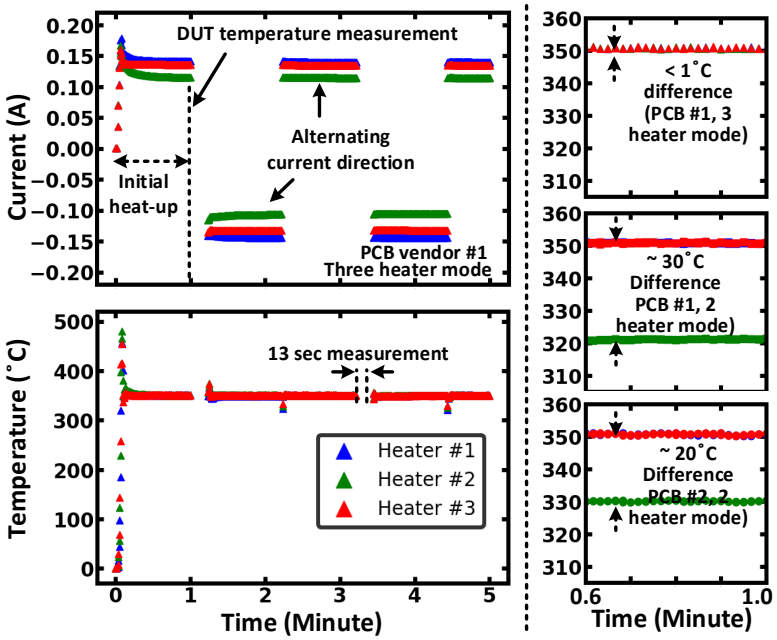


Fig. 3.9. (Upper left) Heater current control for the three heater mode. (Lower left) Heater temperatures for the three heater mode. (Right) Heater temperatures for the different modes

The die temperature is configured by applying high currents to each heater. As illustrated in Fig. 3.8 and Fig. 3.9, a software-based automated control loop was developed to reach the target temperature during the initial heat-up phase. After that, the metal grid temperature is measured, and the current direction is periodically toggled to prevent EM in the heaters themselves. Thanks to the accurate heater control capability, three different temperature gradient conditions were attained, as displayed in Fig. 3.9 (right).

3.3.2 EM Failure Measurement

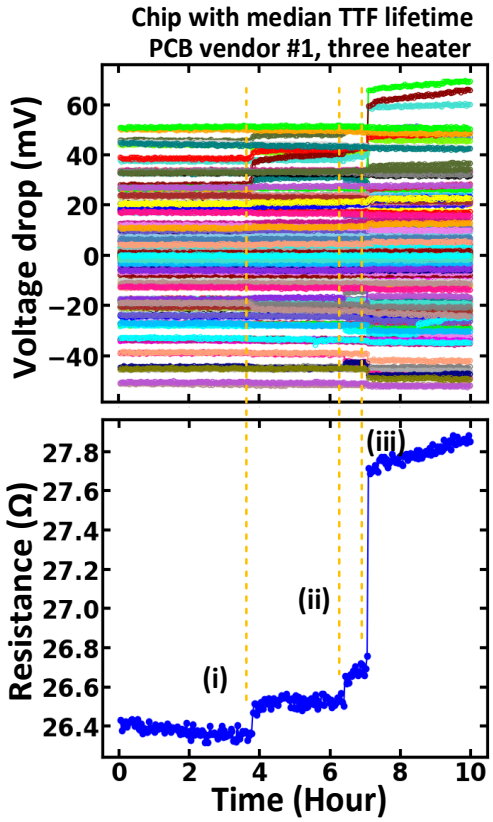


Fig. 3.10. EM data. (Upper) 192 voltage traces and (lower) DUT resistance versus stress time.

After the temperature become stabilized, a 10mA EM stress current is applied to the metal grid. At the same time, 162 voltages of the internal tapping nodes are recorded every 2 minutes. When a voltage shift greater than 10% is detected, the time and location of the EM failures are recorded for data analysis. For example, as shown in Fig. 3.10, once a void formation inside the grid results in changes in the voltage drop profiles, the first (i), second (ii), and third (iii) time-to-failures (TTF) and the node numbers with maximum abrupt shifts are recorded.

3.4 EM Failure Results and Discussions

3.4.1 Time to Failure Statistics

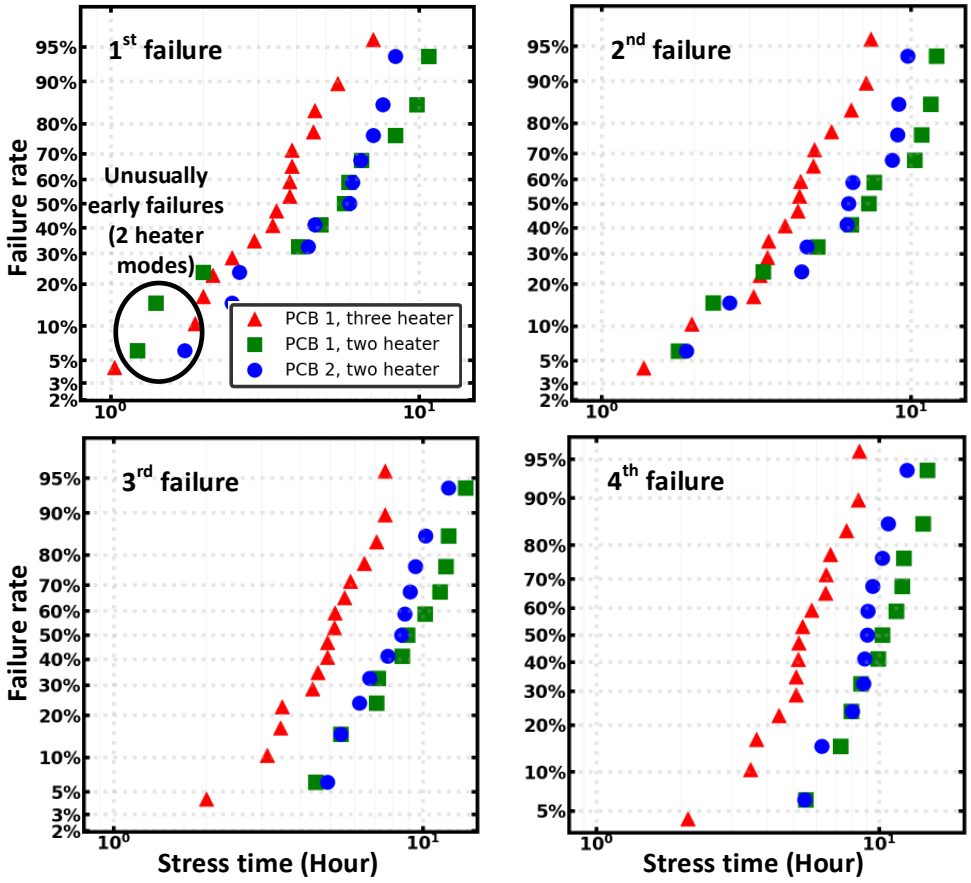


Fig. 3.11. Time to failure distributions for first, second, third, and fourth failures of each test condition. The two heater mode in green shows a bimodal behavior owing to the temperature gradient. For subsequent failures, the distributions revert to the standard lognormal shapes.

Fig. 3.11 compares the cumulative time-to-failure (TTF) plots of the 1st to 4th failure of each test mode. In the 1st failure trend, the uniform temperature condition is close to the typical lognormal distribution, while the other nonuniform modes have bimodal distributions. The deviation from a typical lognormal failure distribution increases when the gradient increases from 0, 20, and 30°C. Furthermore, a comparison between the three heater (red) and the two heater cases (green and blue) suggests that a larger thermal gradient is likely to be responsible for the unusually early failures, as experimented in [15]. A subtle but important observation here is that the heater boundary regions with high temperature gradients are susceptible to EM voidings, as shown in the

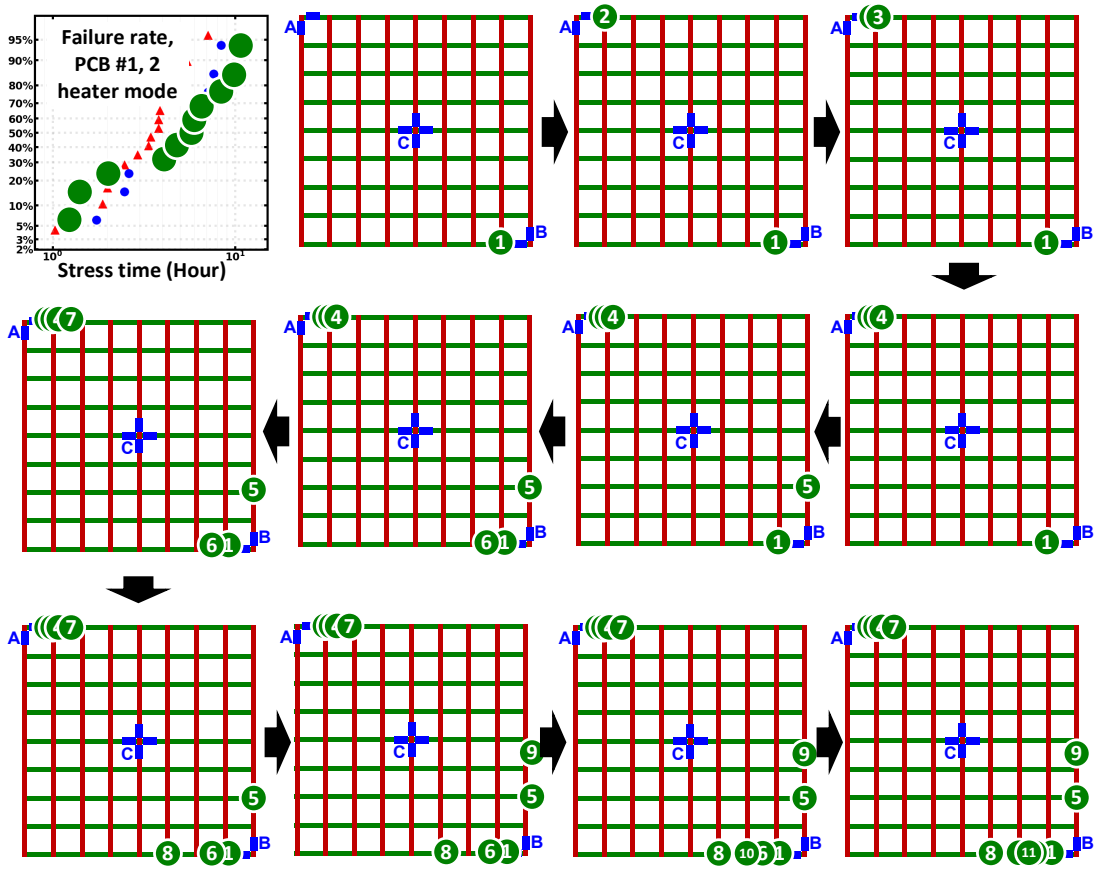


Fig. 3.12. (Upper left) Time to failure distribution of PCB #1, 2 heater mode highlighted in green. First EM void location of each of the 11 chips are shown in the order at which they occurred (e.g., first chip to fail denoted as '1', etc.).

TTF order in Fig. 3.12. Note that the TTF trends return to the typical lognormal distribution by the 3rd and 4th failure. We suspect this is due to the stress currents being redirected away from the earlier failure locations, which may have occurred in the temperature gradient regions.

3.4.2 Failure Locations

Fig 3.13 shows the first failure location of all 38 chips for each test condition. Detailed dimensions of the DUT and the heaters are given in Fig. 3.14 to show the exact location of the metal grid with respect to the three heater stripes. The boundaries of the heaters are located right next to via #18 (upper left corner) and via #62 (lower right corner). Our unique test structure allows the temperature gradient to occur near the heater boundaries which is attained by simply turning off the middle heater. This thermal gradient explains the different failure locations in Fig. 3.13 (left) and Fig. 3.13 (middle,

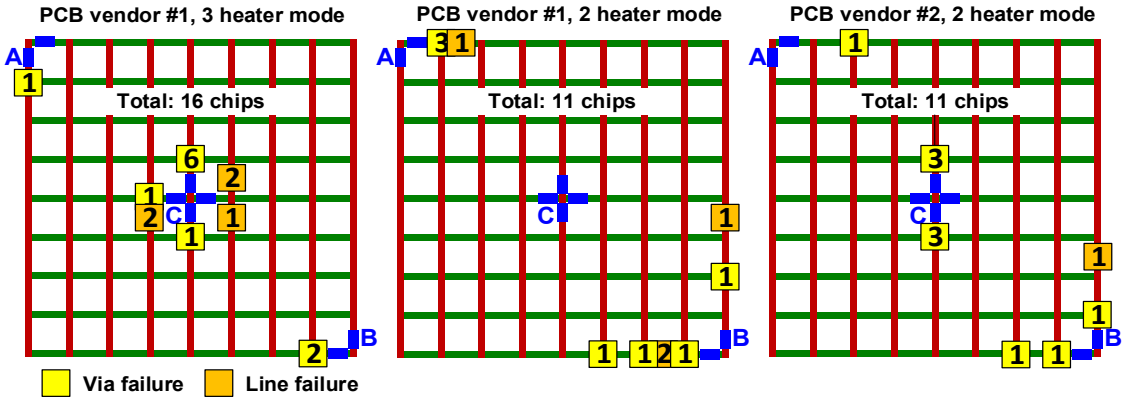


Fig. 3.13. Location and occurrences of the first EM failures for each test condition. Early failures can be seen in the center figure along the bottom horizontal wire where the temperature gradient is the highest.

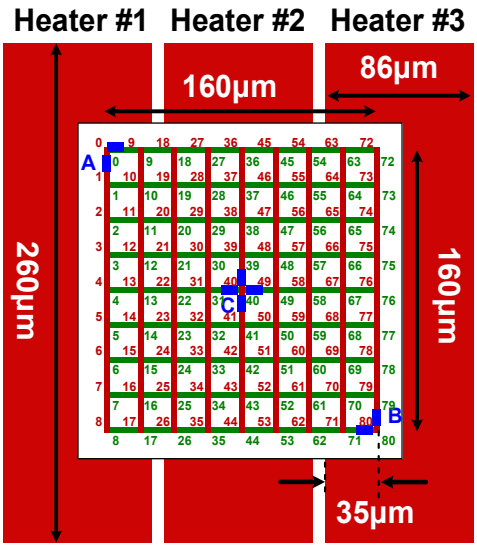


Fig. 3.14. On-chip heater and power grid layout and dimensions, drawn to scale.

right). For the uniform temperature case, EM events are concentrated near the center due to the high current density and high temperature. However, as seen in Fig. 3.13 and Fig 3.15, for the 30°C and 20°C gradient cases, failures occur not only in the high temperature regions but also at locations with a high thermal gradient (e.g., node 62). Moreover, for the 30°C gradient mode, failures at the heater boundaries become more frequent. These results suggest that the additional divergence of the Cu atomic flux

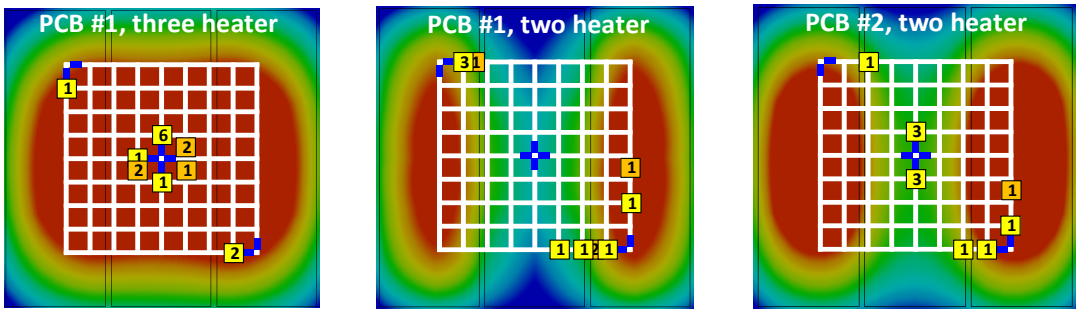


Fig. 3.15. Simulated temperature map overlaid with the test structure showing the impact of temperature and temperature gradient on EM location.

caused by the thermal gradient is responsible for creating additional tensile stress, which accelerates the void nucleation inside the wire.

3.4.3 Detailed Analysis and Comparison with EM Stress Simulator

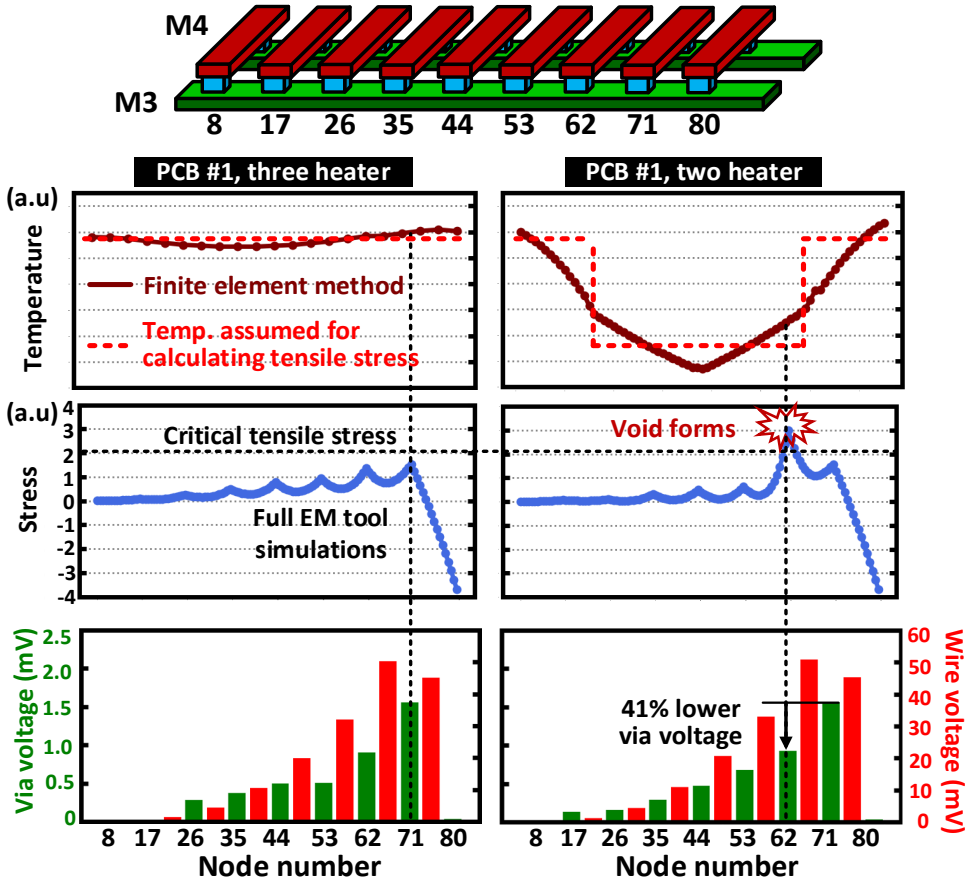


Fig. 3.16. Comparison between three and two heater modes in terms of temperature gradient (from FEM tool), tensile stress (from EM tool), and via/wire voltage drops (measured). Tensile stress in node 62 increases beyond the critical level in the 2 heater mode, causing an EM failure despite the lower voltage and lower temperature.

Fig. 3.16 compares the detailed voltage stress and simulated tensile stress profiles along the bottom M3 segment of the grid for the uniform and nonuniform temperature cases. An industry-grade EM assessment tool that can capture the detailed thermo-mechanical interactions including random distributions of atomic diffusivities and critical stresses, was used to corroborate the experimental data [16]. As seen for both cases, the influx of electrons from each via causes the tensile stress to peak at each via intersection. Stress polarity switched to compressive at the end of the wire due to material confinement. The height of the peak gradually increases along the wire due to the increasing via current. Interestingly, when we apply a temperature gradient along the wire, the tensile stress rapidly increases beyond the critical level due to the phenomena described in Fig. 3.2 – i.e., due to the difference in temperatures, the atomic flux at the anode side is larger than that at the cathode side, generating extra tensile stress as seen in the simulation results. This effect is maximized near the heater boundary (node 62) with the highest temperature gradient, which increases the probability of void nucleation. In this case, even though the EM stress current and temperature are significantly lower, the simulation result clearly proves that the temperature difference between the heater regions accelerates EM void formation. Simulation results from a state-of-the-art EM tool corroborate our findings that thermal nonuniformity in real chips can result in unexpected early failures near temperature boundaries.

Summary

In this chapter, a test chip designed for EM behavior in a power-grid-like 9x9 metal structure is proposed. The focus of this EM experiment is the grid's EM aging trend under different temperature gradients. By using the three on-chip heaters and different test boards, EM stress is done under three conditions with different temperature gradients. The 16 DUTs tested in uniform temperature mode showed typical void nucleation behavior under the high current density regions. However, the 20°C and 30°C temperature gradient data (11 DUTs for each data) showed failures near the maximum temperature gradient regions, even showing earlier failures than the uniform temperature condition. The physics-based EM-induced stress evolution simulator validates that the high temperature gradient increases the atomic flux divergence due to the different atomic diffusivity between the hot spots and cooler regions.

Chapter 4 Electromigration Experiment From Realistic Power Grid Structure

4.1 Introduction

Electromigration (EM) in power grids is a critical reliability concern due to high direct current and Joule heating that results in excessive IR drop. However, characterizing a power delivery network's EM behavior is not trivial since the grid failure has to be accelerated by a large stress current and extreme (also accurate) DUT temperature (e.g., > 300°C). Another difficulty is monitoring a power delivery from power pads to each logic cell in the grid, which requires special voltage scanning schemes. Thus, previous works have analyzed EM effects in metal grids but were limited to simple test structures [12] or relied entirely on simulation models [17].

In this work, we present a test-chip-based novel end-to-end power grid EM characterization methodology. The EM test vehicle consists of four different grid types with 1024 voltage tapping points per grid for measuring the power delivery status. The voltage measurement circuit is capable of tracking IR drops in power rings, power rails, and quasi-cells. An accurate DUT temperature control loop with a maximum temperature error of 2-3°C was implemented using on-chip metal heaters and a DUT temperature sensor [18]. The grids are tested with different stress currents and temperatures to analyze the time-to-failure (TTF) statistics. Based on the accelerated TTF, the lifetime of the grids is calculated to estimate the reliability of power networks in realistic die temperature and current consumption scenarios. Structure-dependent degradation

behaviors are compared since we have four grid structures with different via and rail strengths. The comparison between the silicon data and an industry-grade EM mechanical stress simulator explains such failure trends with void-forming locations and statistics.

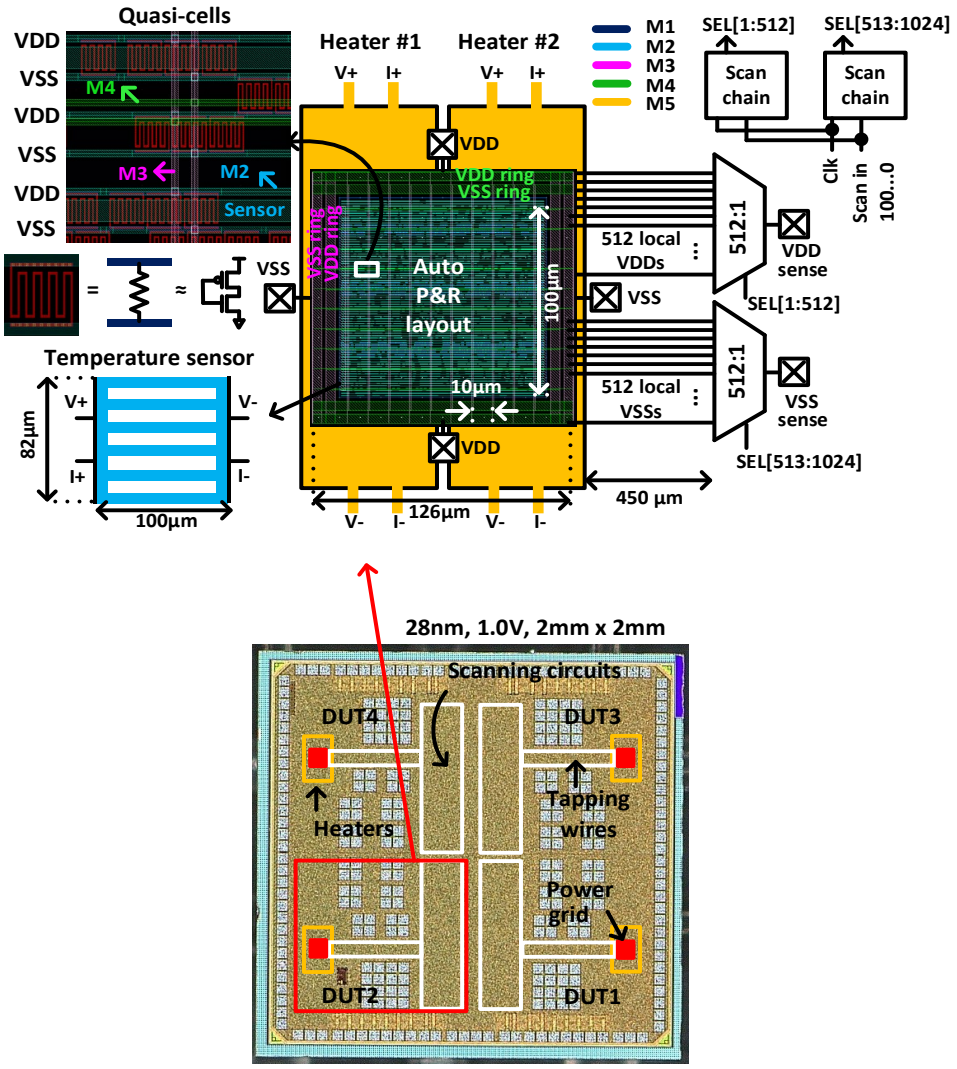


Fig. 4.1. 28nm EM test chip overview. DUT includes a power grid (M2-M4) and quasi-cells with gate equivalent poly resistances. The heating area is covered by two metal heaters (M5) and a temperature sensor (M2). The voltage measurement circuit includes two scan-chains and two 512:1 analog muxes.

4.2 Test Chip Description

An overview of the 28nm power grid EM test vehicle is given in Fig. 4.1. We first synthesized a digital functional block and generated a 100µm x 100µm layout using an automatic place-and-route tool. Then, we replaced the individual logic gate cells with equivalent poly resistor cells so that the load circuits could withstand the stress temperature required for EM acceleration. The automatically generated power grid consists of three metal layers (M2-M4), including power rings and rails, with poly-based quasi-cells. Two upper rails (M3, M4) are connected to the power rings, delivering power to the lower M2 rails and quasi-cells. The power grids were placed inside the heating area, whose temperature is controlled by two on-chip metal heaters (M5) with individual control. A dedicated M2 metal temperature sensor was used for the temperature feedback.

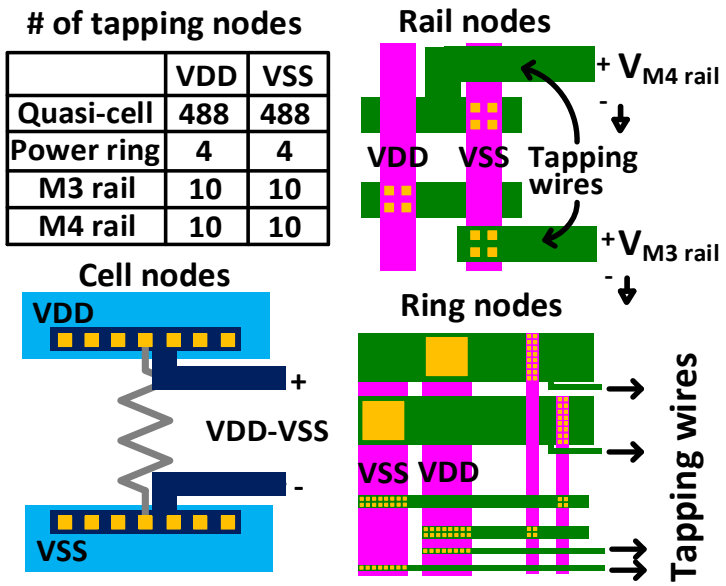


Fig. 4.2. Breakdown of voltage tapping nodes and tapping locations in quasi-cells, power rings, and power rails. The tapping nodes allow measuring voltage delivery from rings and rails to each cell.

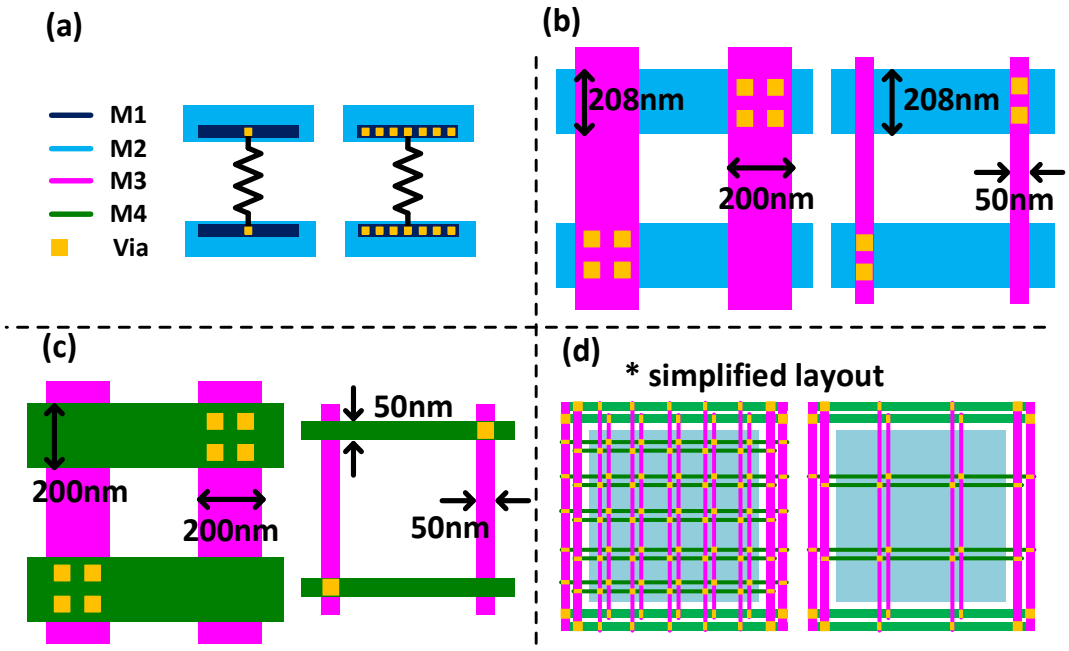


Fig. 4.3. Designed power grids' structural difference details. (a) Cell via counts. Min vs. max cell via (b) M2, M3 rail widths and M2-M3 vias. (c) M3, M4 rail widths and M3-M4 vias. (d) Number of M3 and M4 rails (rail density)

Power grid		DUT1	DUT2	DUT3	DUT4
# of cell via		1	7	7	7
M3, M4 rail width (nm)		200	200	50	50
# of rails	M3 VDD	10	10	10	2
	M3 VSS	10	10	10	2
	M4 VDD	10	10	10	2
	M4 VSS	10	10	10	2

Fig. 4.4. Comparison table of four different DUTs. DUT2 is the baseline design with maximum via and rail strengths.

512 local VDD and 512 local VSS voltage are tapped from across the power grid and routed to two 512:1 analog multiplexers which are controlled by a 1024-stage scan

chain. 488 of the 512 voltages were measured from the quasi-load cells, while the remaining 24 voltages were measured from various points of the power rails and power rings (Fig. 4.2). As illustrated in Fig. 4.3, the four power grid types have different cell via counts, rail widths, and rail densities. The DUT2 is a baseline design with the strongest vias and rails, as shown in Fig. 4.4.

4.3 Experimental Methodology

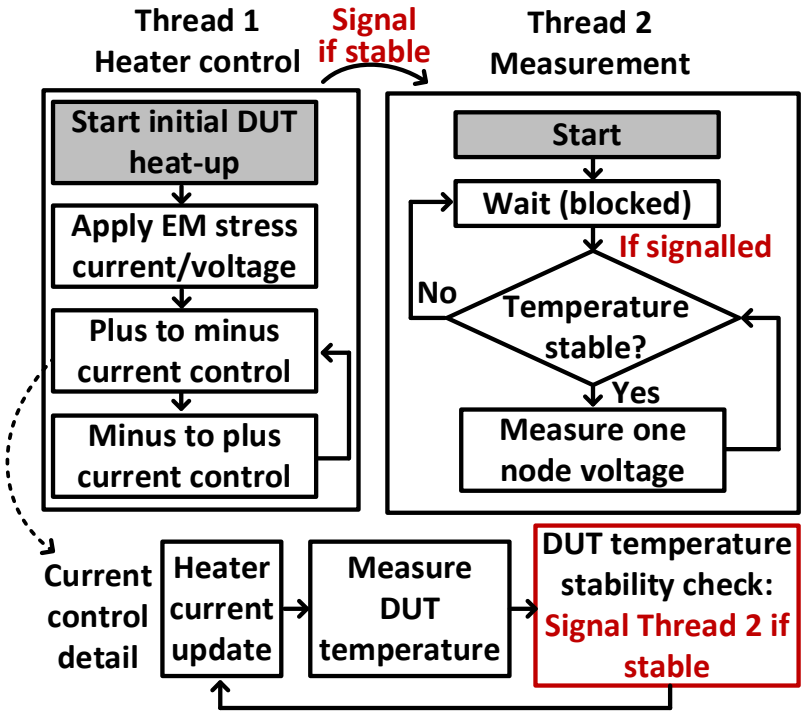


Fig. 4.5. Experiment flow with two concurrent controls. Thread 1 maintains consistent DUT temperature. Thread 2 runs only if the DUT temperature is fully stable.

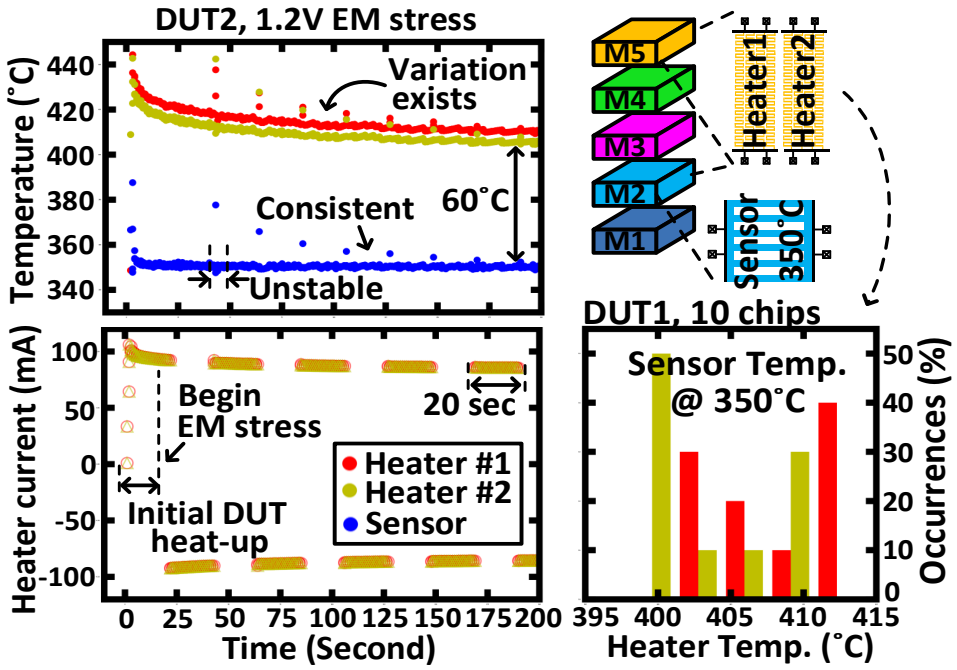


Fig. 4.6. (Top left) Temperature control logs and heater current data. (Bottom left) Measured heater currents. (Right) Histogram of heater temperature under 350˚C sensor temperature.

Fig 4.5 shows the experiment flow, including the heater control loop and voltage measurement loop running concurrently. The accuracy of the DUT temperature is critical because the grid resistance and the IR drop profile can fluctuate with inconsistent grid temperature. To this end, a multi-threaded control software performs the voltage measurement only when the temperature is stable. The measured sensor and heater temperatures, along with the heater current, are shown in Fig. 4.6.

The target DUT temperature here is 350°C (±1°C). The temperature coefficient of resistance (TCR) method was used to translate the heater and sensor resistances to the corresponding temperatures (Fig. 4.7). Due to the temperature gradient from M5 (heater) to M2 (sensor), the sensor temperature was ~60°C lower than the heater temperature

under a steady-state stress condition. Once the sensor temperature reaches the target, which is within 10-20 seconds, a constant stress current or stress voltage is applied. To avoid EM damage in the metal heaters themselves, the direction of the heater current is switched every 20 seconds. As illustrated in Fig. 4.6 (right) histogram, the DUT temperature is accurate at 350°C even though the heater temperatures of each chip are different.

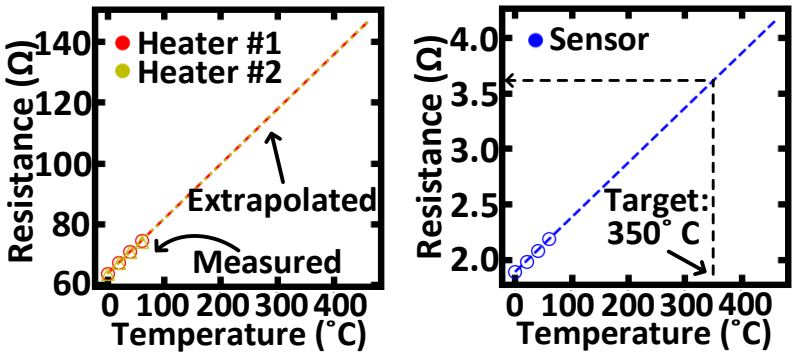


Fig. 4.7. Temperature coefficient of resistance (TCR) of the heaters and the temperature sensor

4.4 Power Grid Failure Data Analysis Methodologies and Lifetime Prediction

Fig. 4.8 (a) is a measured fresh chip voltage map of 488 cell VDDs and VSSs. By subtracting the VDD and VSS of each cell, the power delivery status is visualized (Fig. 4.8 (b)). Note that VDD-VSS values are higher at the edge due to the IR drop. The voltage shift of VDD-VSS was calculated using the initial fresh-state voltage maps, as illustrated in Fig. 4.10 error maps, to monitor the aggravated IR drops after the EM stress.

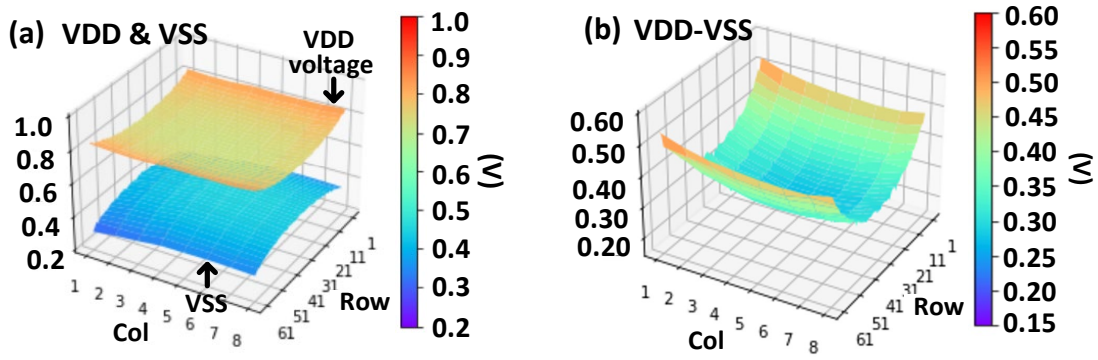


Fig. 4.8. (a) Measured fresh chip DUT2 cell VDD & VSS voltage map. (b) Time zero cell VDD-VSS voltage map.

4.4.1 Current vs. Voltage Stress

Fig. 4.9 explains how the TTF was measured and the EM beginning point was captured. As illustrated in Fig. 4.9 (a), the grid resistance from the VDD power ring to the VSS power ring (i.e., pad resistances are excluded) increases gradually due to the growth

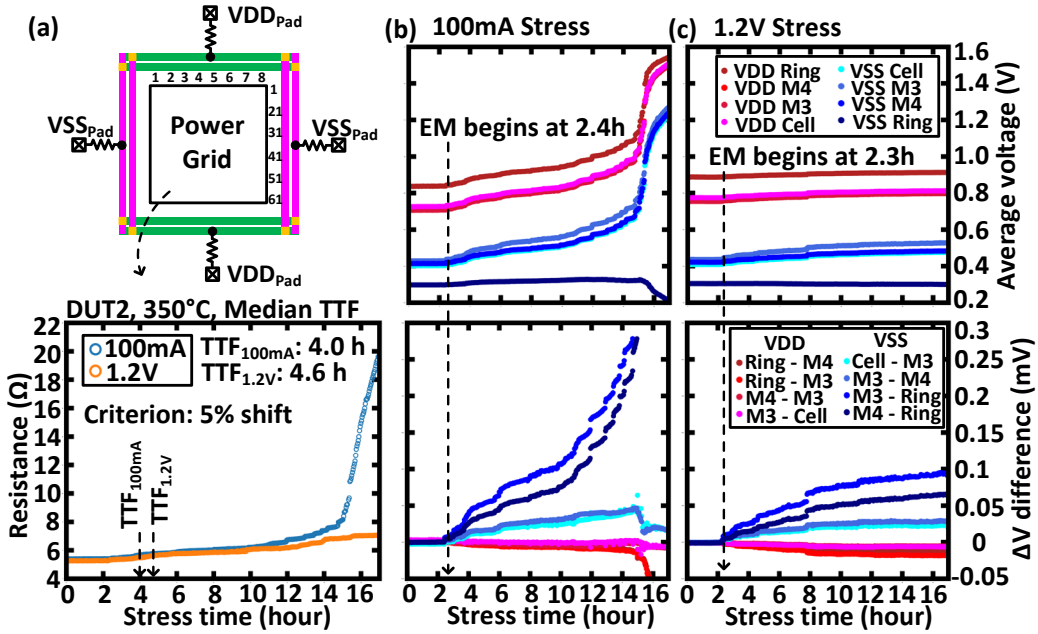


Fig. 4.9. (a) Power grid resistance shifting profiles are plotted for TTF measurement. The TTF criterion is a 5% resistance shift. (b-c) Comparison of current and voltage stress. The ring, rail, and cell voltages are averaged (top). The voltage differences between the rings, rails, and cells for capturing subtle EM beginning times and locations (bottom).

of EM voids in the grid. We defined the TTF criterion as a 5% resistance shift. Also, the ring, rail, and cell voltages are averaged and plotted to capture the subtle EM-induced local resistance shift, as shown in Fig. 4.9 (b-c). For instance, if the 1.2V voltage is applied to the power pads, the EM begins at 2.3h, while the $TTF_{1.2V}$ is 4.6h. For DUT2, the 1.2V pad voltage is equivalent to 100mA total current at 350°C, and we both tested the constant voltage and current modes. The grid degradation of the voltage mode (Fig. 4.9 (c)) is slower than the current mode (Fig. 4.9 (b)) since its stress current becomes smaller with the increase of the grid resistance.

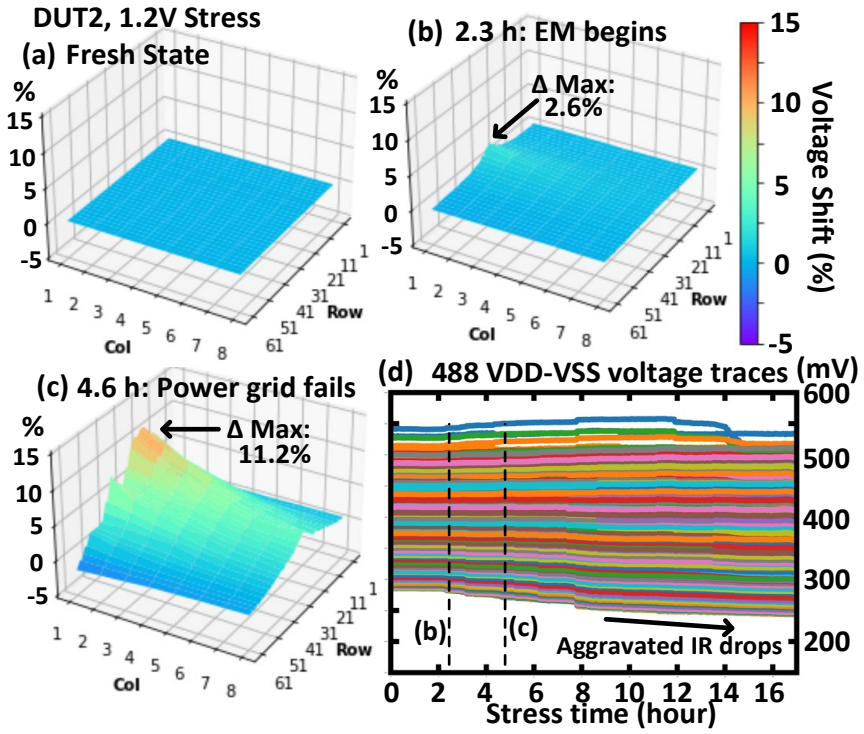


Fig. 4.10. Error map (voltage shift) of each cell VDD-VSS at (a) fresh state, (b) beginning of EM, and (c) when the power grid fails (TTF). (d) The power delivery is gradually aggravated after the EM stress.

DUT2, 350°C temp.

Stress mode	Average TTF (hour)	# of chips
1.2V	4.3	5
100mA	4.0	13

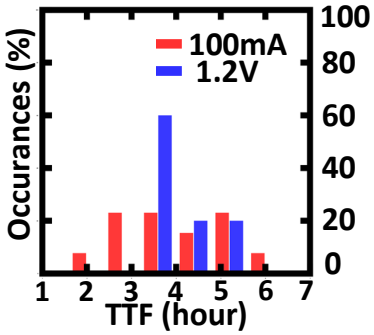


Fig. 4.11. Current & voltage stress average TTF (left) and TTF histogram (right)

The error map of VDD-VSS in Fig. 4.10 suggests that the excessive IR drop is relatively small at the beginning of EM (2.6% error) but fails completely after TTF (>10% error). The two modes have similar TTFs (Fig. 4.11) because they have almost equal grid currents.

Fig. 4.12 is a statistic of the time zero DUT resistance at 350°C. It is well shown that DUT2 is the strongest grid with the lowest resistance, followed by DUT1, which has fewer cell via count. DUT3 and DUT4 have high resistance since they have narrower rails (Fig. 4.3 (b-c)) and fewer rails (Fig. 4.3 (d)), respectively.

Res. measurement at 350°C

Grid type	Avg. (Ω)	Std.	# of DUTs
DUT1	5.5	0.09	10
DUT2	5.3	0.13	33
DUT3	7.0	0.18	9
DUT4	29.8	0.64	4

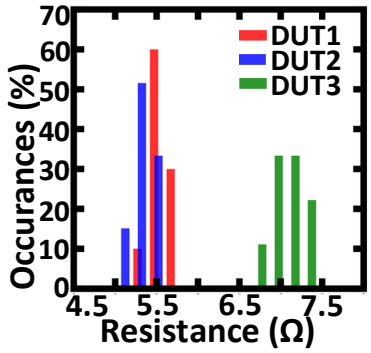


Fig. 4.12. Time zero grid resistances at 350°C. (Left) Average resistances. (Right) Histogram of the resistance

4.4.2 Temperature and Current Dependent Lifetime Behavior

Thanks to the accurate DUT temperature control, the EM experiment was done under 375°C, 350°C, 325°C, and 300°C modes. For a better understanding, we assumed that the lifetime of the grid tends to, although not strictly, follow the temperature and current density dependency of a widely used Black's equation (expression (2)).

As shown in the grid resistance shift in Fig. 4.13 (a), the TTF increases rapidly with the linearly decreasing temperature. Similarly, reducing the stress current from 120mA to 60mA improves the grid lifetime (Fig. 4.13 (b)). The measured average TTF in Fig. 4.14 and the histograms in Fig. 4.15 clearly show the trend.

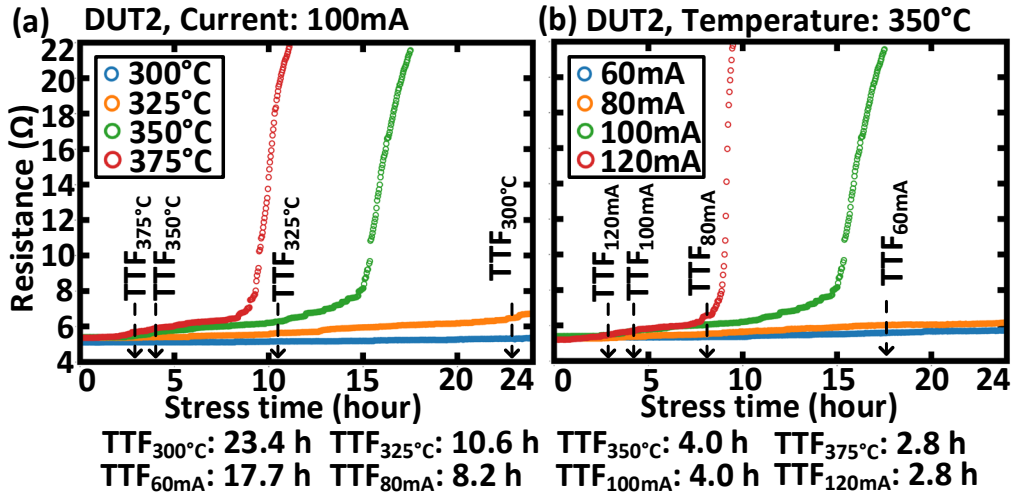


Fig. 4.13. Power grid resistances and failure time comparison under (a) four temperatures and (b) four currents, respectively.

DUT2, 100mA current

DUT temp. (°C)	Average TTF (hour)	# of chips
300˚C	27.0	5
325˚C	9.9	5
350˚C	4.0	13
375˚C	2.9	5

DUT2, 350°C temp.

Stress current (mA)	Average TTF (hour)	# of chips
60mA	15.7	5
80mA	7.5	5
100mA	4.0	13
120mA	2.8	5

Fig. 4.14. Measured average TTFs. (Left) temperature dependent TTF. (Right) current dependent TTF

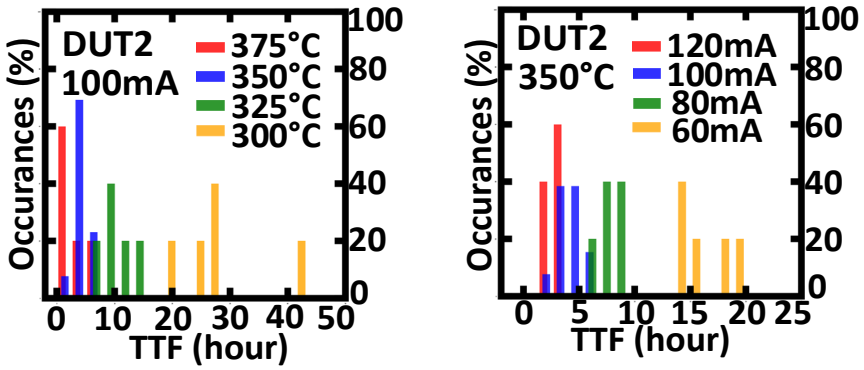
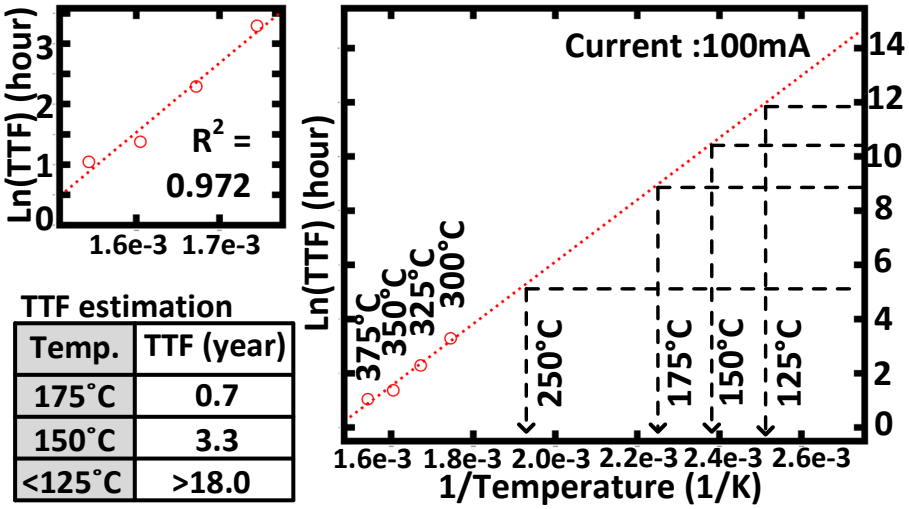


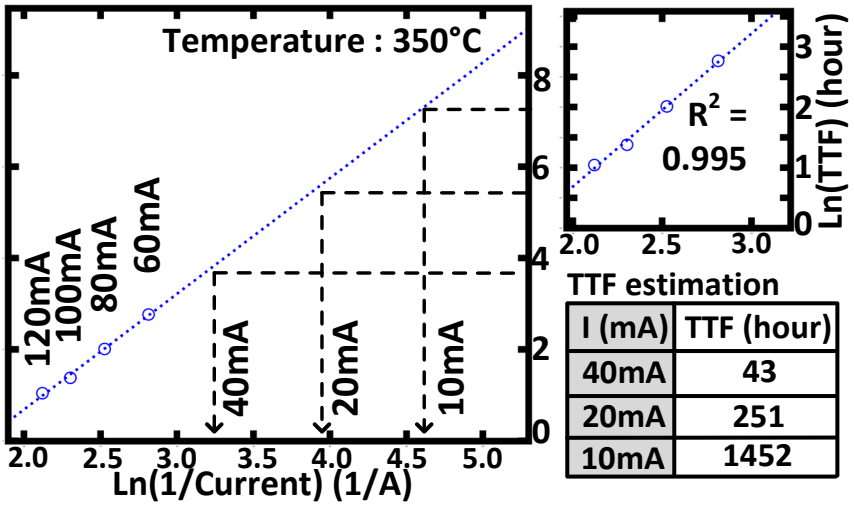
Fig. 4.15. TTF histograms. (Left) temperature dependent TTF. (Right) current dependent TTF

Fig. 4.16 is a log scale plot of the average TTF for grid lifetime estimations. The measured average TTFs are linearly extrapolated to the lower temperature and current regions.



TTF estimation

Temp.	TTF (year)
175˚C	0.7
150˚C	3.3
<125˚C	>18.0



TTF estimation

I (mA)	TTF (hour)
40mA	43
20mA	251
10mA	1452

$$MTTF = \frac{A}{j^n} \cdot \exp\left(\frac{E_a}{k \cdot T}\right)$$

Fig. 4.16. Average TTFs are extrapolated for grid lifetime estimation. Black's equation (bottom) is assumed for the TTF calculation just for the simplicity of the lifetime analysis.

The R^2 value of the linear trend line is 0.972 and 0.995, respectively. The TTF estimation table shows that the DUT2 grid's lifetime below 125°C is over 18 years. Similarly, the calculated TTF with a 10mA grid current is 1452 hours. Note that the model used for the lifetime predictions has limited accuracy for grids with many redundancies since Black's equation-based lifetime estimation is normally for single wire interconnects. Further studies are required for the grid EM TTF model calibrations for complicated power nets.

4.5 Geometry-Dependent Failure Analysis and Model Calibration

1.2V stress, 350°C

Grid type	Average TTF (hour)	# of chips
DUT2	4.3	5
DUT3	2.2	4

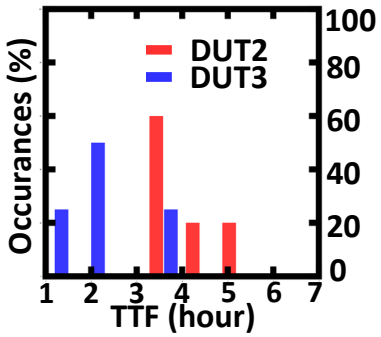


Fig. 4.17. DUT2&DUT3 TTF. Average TTF (left). TTF histogram (right)

The reliability of a power grid highly depends on its via counts, rail widths, and rail densities. Fig. 4.17 is a comparison of the measured average TTF of the DUT2 (strong grid) and the DUT3 (narrower grid) with 1.2V constant voltage stress mode.

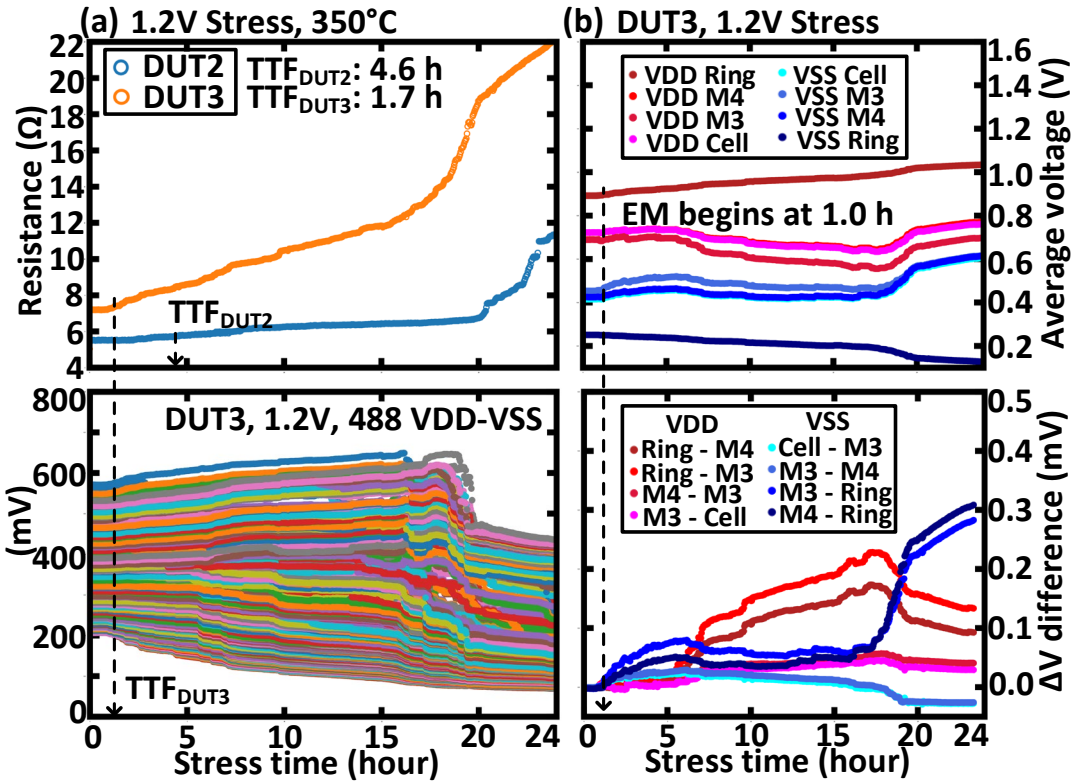


Fig. 4.18. (a) TTF comparison of DUT3 and DUT2 (top) and DUT3's cell voltage profile (bottom). (b) Rings, rails, and cell voltage traces of DUT3

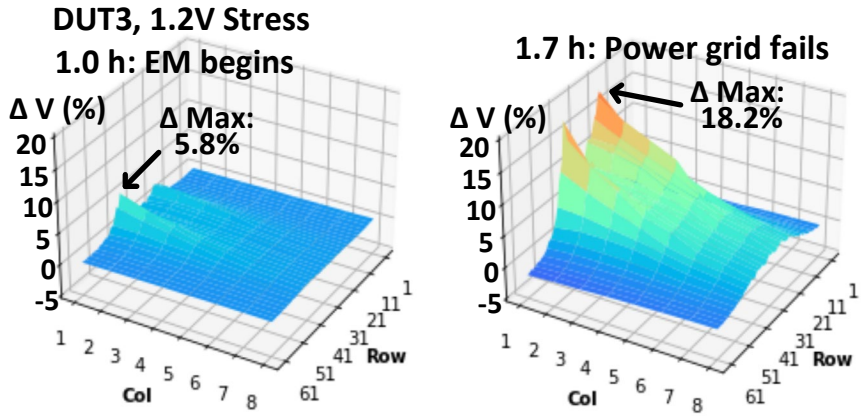


Fig. 4.19. Error maps of DUT3 (narrower rails)

As shown in Fig. 4.18, the narrower rail results in an earlier EM with smaller TTFs. Also, due to the worse excessive IR drop, DUT3 reaches an 18% cell VDD-VSS error rate at its failure time (Fig. 4.19).

DUT1-3: 100mA
DUT4: 10mA, 350°C Temp.

Grid type	Average TTF (hour)	# of chips
DUT1	4.1	10
DUT2	4.0	13
DUT3	1.3	5
DUT4	3.8	4

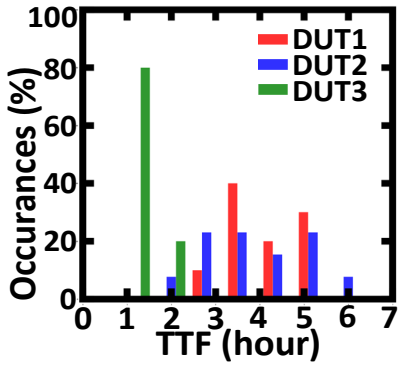


Fig. 4.20. Measured TTFs of different types of grids

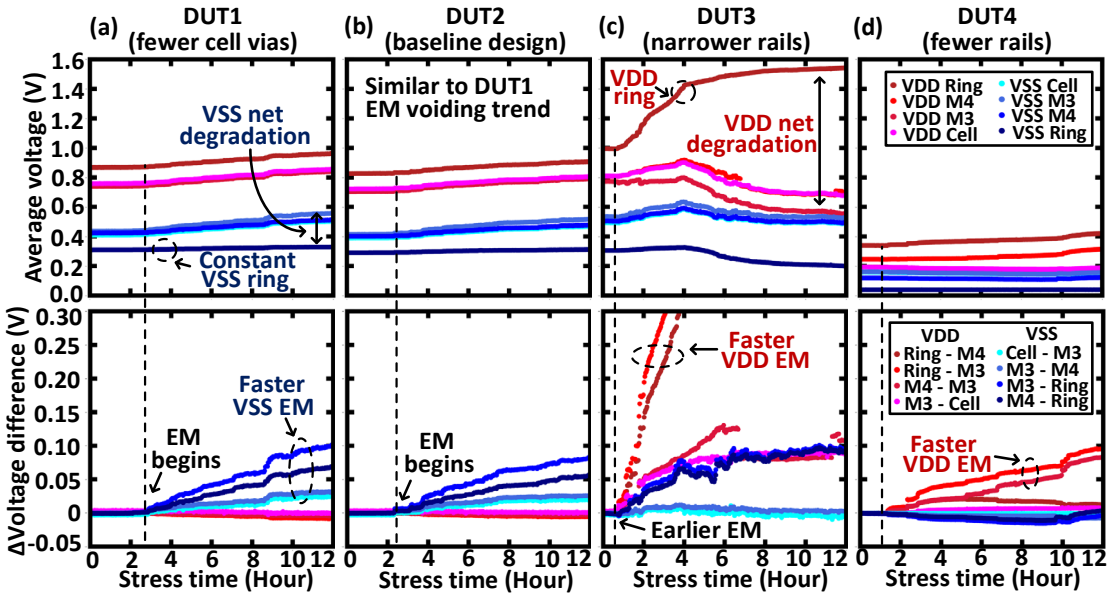


Fig. 4.21. Measured voltage traces at 350˚C. (a-c) 100mA stress current. (d) 10mA stress current.

For physics-based EM simulator model calibrations [2][3], the four DUTs are tested with constant stress currents. Note that a smaller 10mA current is used for DUT4 since the grid is significantly weaker than the others. Fig. 4.20 shows that DUT2's TTF is longer than DUT3, as expected from the previous voltage stress mode. However, there is no clear lifetime difference between DUT1 (minimum cell via count) and DUT2 (maximum cell via) (Fig. 4.3 (a)). The detailed trend is shown in Fig. 4.21 average voltage plots. As seen in the data, the voltage traces of DUT1 and DUT2 are very similar, which suggests that the number of vias inside the quasi-load cell doesn't play a major role in the EM voiding trend. This is because the current density is much higher in the metal grid compared to that in each quasi-load cell. The DUT1 and DUT2's voltage shifts in the VSS rails are greater than those in the VDD grid. On the other hand, DUT3 and DUT4 also show significant voltage changes in the VDD nets. This suggests that the voids can hurt the power delivery not just locally but also in the multiple power nets if the rail widths and rail spacings are poorly chosen.

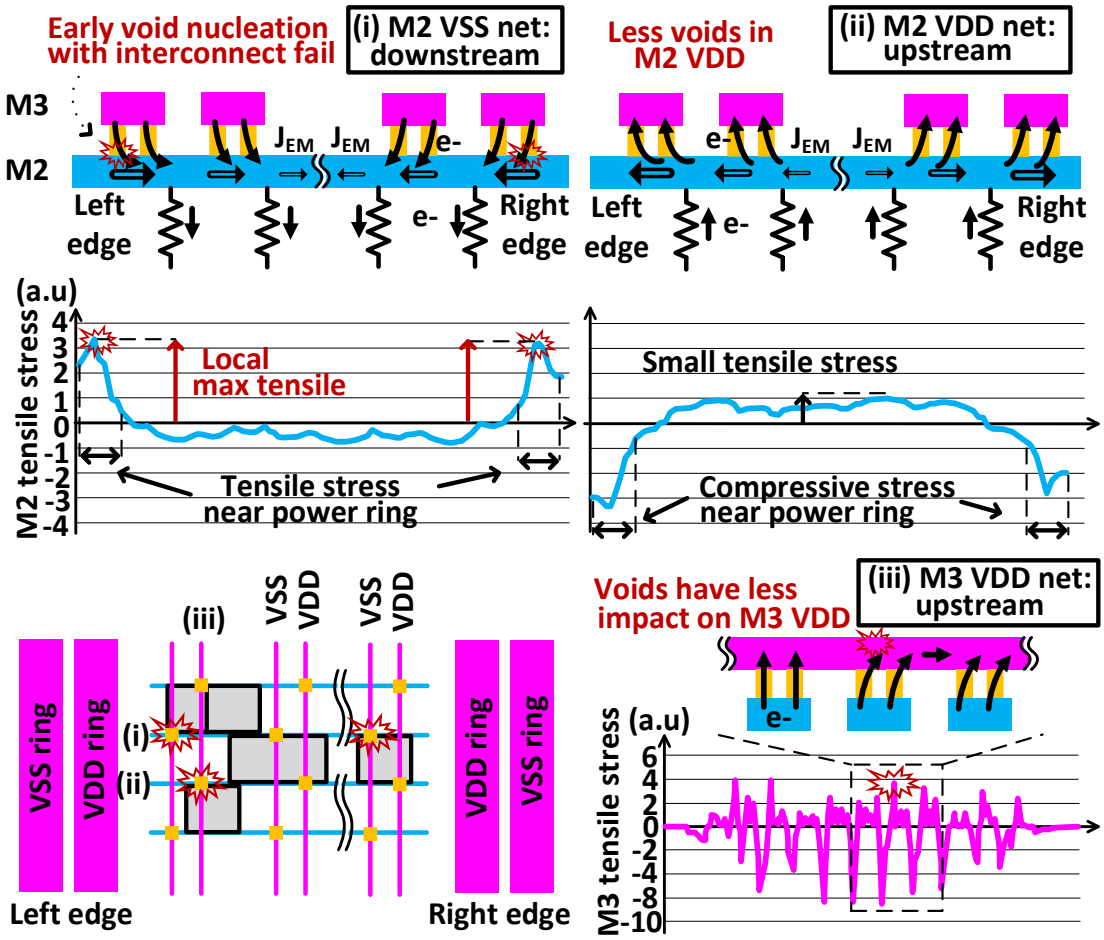


Fig. 4.22. EM void simulation results for M2 and M3 rails.

The extensive test data was used to calibrate the parameters of an industry-grade physics-based EM simulator (Fig. 4.22). Since the downstream scenario in the M2 VSS net results in high tensile strength near the power ring regions, the early void nucleations occur on the left and right edges of the M2 VSS rails, especially under the vias, causing abrupt IR drops. Note that the EM-induced atomic flux is maximized at the left and right edges of the M2 rails. Since the power delivery starts from the power ring (i.e., the left and right edge of the M2 rails), electron wind is minimized in the middle of the power

rails due to the escape of the electrons to the resistive loads. Therefore, in the M2 VDD net, the compressive stress is maximized at the edge regions, and the middle region has lower tensile stress due to the lower current density from the current distribution. Furthermore, in the VDD nets, high tensile stress is not formed inside the M2 rails but in M3. Such above-via voids may have less impact on IR drop, which can explain the asymmetry in the VDD and VSS degradations (Fig 4.23). Also, the number of voids predicted by the EM tool in Fig. 4.24 explains why EM is more dominant on the VSS side.

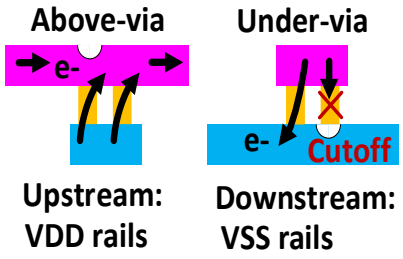


Fig. 4.23. Voiding locations in upstream/downstream EM scenarios.

DUT1 void simulation

Stress direction	# of voids (normalized)
M3 to M4	3.0
M2 to M3	2.3
M4 to M3	1.0
M3 to M2	4.7

Fig. 4.24. The number of voids predicted by an industry-grade EM void simulator validates the earlier EM in DUT1 VSS nets.

Summary

In this chapter, a test chip for EM failure characterization in an on-chip power delivery network is proposed. The purpose of the test chip is to compare the power grid's geometry-dependent, grid current-dependent, and temperature-dependent EM lifetime behavior. The on-chip voltage measurement circuitry allowed the scanning of the internal voltage of each cell and rail in the grid. Utilizing the detailed voltage scanning capability, the changed power delivery profiles were analyzed to narrow down the EM voiding locations. By monitoring the grid resistance shift, TTF was calculated and compared. As expected, the power net's lifetime is longer with wider and denser power rails, more via structures, due to lower current density. However, the cell-via redundancy didn't show noticeable aging behavior differences since the current density in each cell is significantly lower than the power nets. As expected, high temperature and high current accelerated the aging of the power delivery network and showed faster IR drops. The EM-induced stress simulator validated that the faster VSS net failure is from high tensile stress in the power nets near the power ring regions. This is because the cathode net shows more material depletion than the rest of the power net. The voiding simulator also proved that the VSS net dominates the grid resistance increase. However, if the grid is poorly designed with narrower rails and less via redundancy, the impact of voids in VDD nets is not negligible due to the high current density in the power networks.

Chapter 5 Advanced On-Chip Heater Control Methodology

5.1 Introduction

Designing an EM test-chip, including measurement circuits, is not trivial due to the high temperature required for EM acceleration. As explained in previous chapters (Chapter 3 and Chapter 4), a localized heat source that only increases the temperature of a DUT, not the measurement circuitry with active devices (i.e., transistors) is important for the EM test chips. This is because the typical EM acceleration conditions (>200°C) can hurt the EM monitoring circuits since the silicon transistors cannot withstand such high temperatures. Therefore, placing the test chip inside an oven setup is not available with this approach. To utilize the advantage of circuit-based silicon reliability data collection, using on-chip heaters is necessary for flexible design and experiments.

5.2 On-Chip Heater Design and Control Overview

There are many on-chip heater-based experiments in various reliability testing approaches, but not many works explained the details of the heater structures and the heater controlling methodologies. H. Yu et al. [18] proposed detailed metal heater structures that worked in various processes (350nm, 65nm, 28nm, 16nm) and showed their general control methodology.

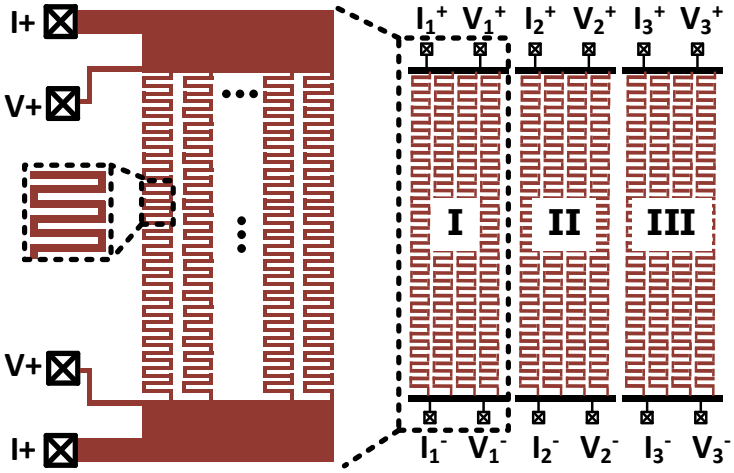


Fig. 5.1. (Left) On-chip heater with four-terminal Kelvin testing connection. (Right) Multiple heaters can be used to raise the temperature of a larger circuit area. [18]

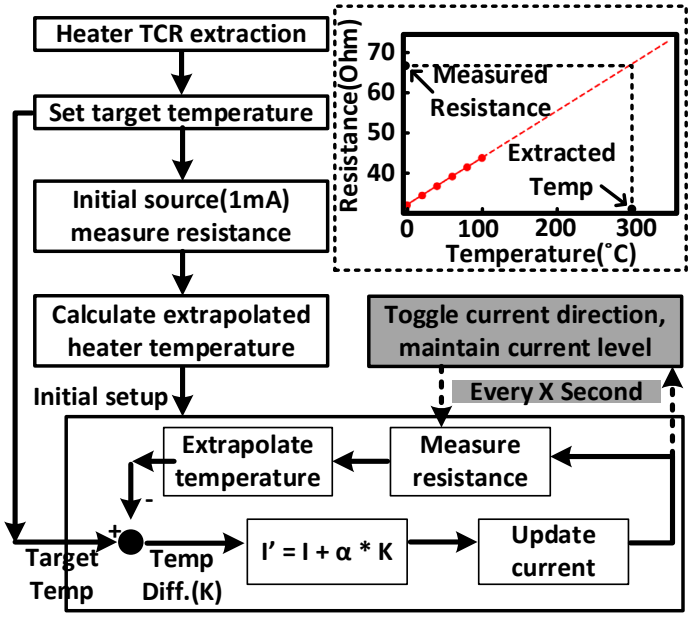


Fig. 5.2. Heater temperature control loop. [18]

Fig. 5.1 is an overview of a metal on-chip heater structure with long snake-shaped wires. The thin, narrow metal wire is used to maximize the Joule heating per unit area to increase the heater temperature effectively in a short amount of time. Each heater has four ports for 4-wire Kelvin sensing for accurate heater resistance measurement. The accuracy of the resistance is critical for the heater temperature control, which is explained in Fig. 5.2.

The very first step of the heater temperature control is measuring the temperature coefficient of resistance (TCR) of each heater. As shown in Fig. 5.2, the heater resistance is measured at several different temperatures under 100°C. If the heater is integrated with measurement circuitries on the same die, generally, the resistance measurement is done at less than 100°C conditions so as not to damage the active devices. The measured

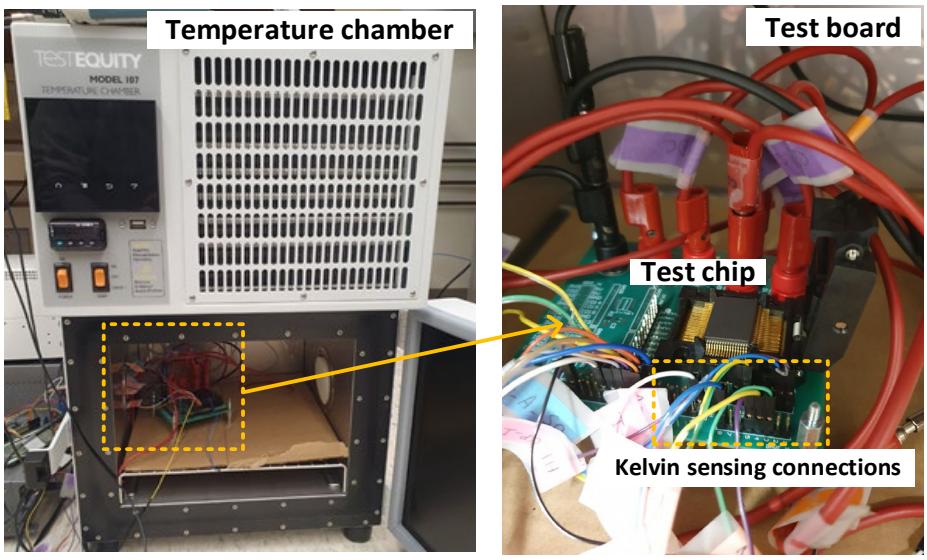


Fig. 5.3. Temperature chamber (left). Board in the chamber with the 22nm EM test chip (metal heaters included) in the socket (right)

resistances are linearly extrapolated to the extreme temperature regions (>300°C) and used as a reference for a temperature translation. If the TCR of a metal heater shows an expected linear trend (e.g., $R^2 > 0.99$), the slope and the y-axis intercept are calculated from the extrapolation, and a resistance measurement can be translated to the temperature value.

Fig. 5.3 is a typical experiment setup for heater TCR measurements. The temperature chamber (Fig. 5.3 (left)) can change the internal temperature from -20 °C to 120 °C. Inside the chamber, the test board, including the EM test chip is located. The 4-wire measurement of the heater is generally done with source meter units (SMU), which enables accurate kelvin sensing.

After the TCR is acquired, the heater control flow illustrated in Fig. 5.2 is done to reach the target heater temperature. The purpose of the feedback loop is to update the heater's current value to maintain the heater or DUT's temperature consistently to the target temperature. For instance, if the target temperature is 300°C, the control software increases/decreases the heater current proportionally to the difference between the 300°C target and the current heater temperature. Once a stable 300°C is achieved, one can start the reliability experiment on a DUT. Note that the heater's current direction changes periodically in the control flow. This is because unidirectional currents in the metal heater can damage the heater wires from EM. Since the alternating current direction can significantly increase the heater lifetime from the EM self-healing effect, switching their direction every 10-20 seconds is recommended.

5.3 Heater Temperature Control vs. DUT Temperature Control

The previous section introduced a general heater temperature control methodology. However, even if the heater covers the DUT area on a layout, the temperature of the heater doesn't guarantee the temperature of the DUTs if they are vertically away from each other. Such vertical distance between the heat source and the DUT often results in non-trivial temperature discrepancy, as shown in Fig. 5.4. In this 28nm test chip example, the heaters are at the M5 metal layer and the M1 DUT is below the M2 temperature sensor. The temperature sensor is simply parallel connected metal

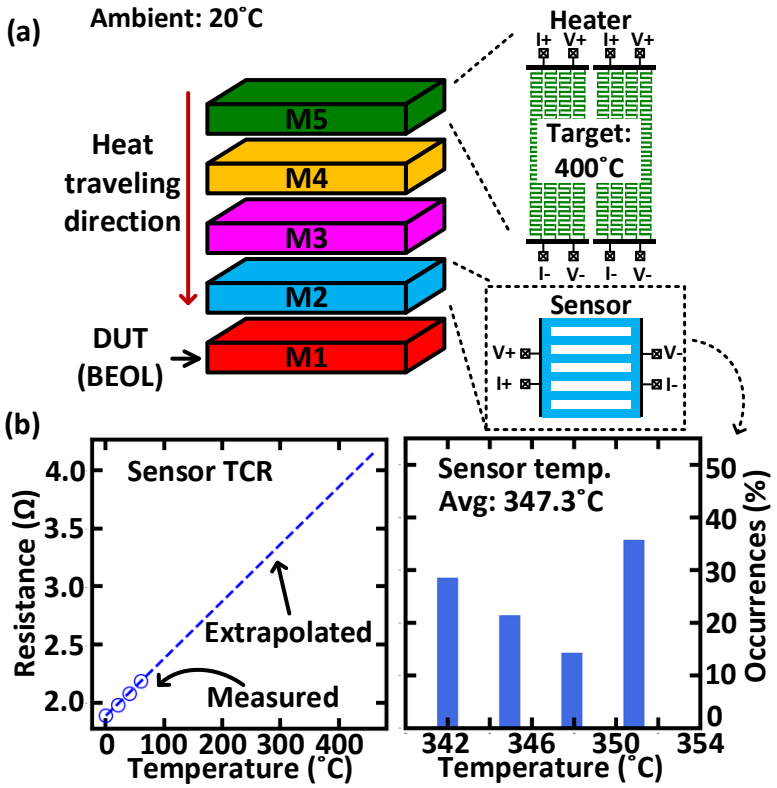


Fig. 5.4. (a) Vertical distance between temperature sensor and metal heaters, (b) Histogram of the sensor temperatures of 14 chips when the heater temperature is regulated at 400°C. 28nm process technology

lines with four port connections also for accurate resistance measurement using Kelvin sensing. Similar to the heater TCR, the sensor's TCR is used to measure the temperature of the DUT since the sensor is closely located near the DUT. As described in the Fig. 5.4, there are around 50°C difference between the heaters and the sensor due to the vertical distance. Also, the sensor temperatures are not consistent for each chip, as shown in Fig. 5.4 ((b) right). The reason for the DUT temperature variation is a variation of thermal conductivities. Since the thermal conductivity of each silicon die is not identical, the amount of heat reached from the heaters to the DUT generally shows variations. In addition to the thermal conductivity difference, the amount of Joule heating from the DUT also contributes to the variation. For example, if the DUT is a power grid for an EM experiment with high power dissipation (e.g.,100mW power consumption), the variation

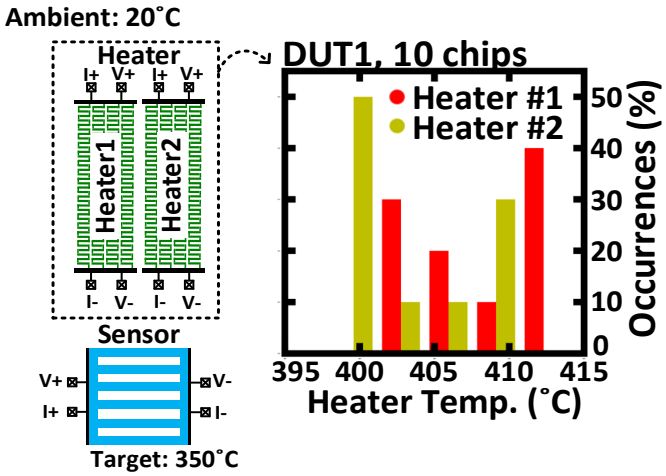


Fig. 5.5. Histogram of the sensor temperatures of 10 chips when the sensor temperature is regulated at 350°C

of DUT resistances can lead to different power dissipation of the DUT. Therefore, the amount of heat generated from the DUT itself becomes a variation source.

Therefore, for a consistent DUT temperature control, a heater control loop based on the DUT temperature measured from the dedicated temperature sensor is required. For example, the target and current temperatures in the Fig 5.2 control loop could be replaced from the heater temperature to the sensor temperature. Fig. 5.5 is a histogram of the heater temperatures with the improved heater regulation with the same 28nm test chip. Even though the temperatures of the heaters have variations, the sensor temperature is consistent for all 10 chips since the temperature regulation target is the sensor. Using this advanced method, the DUT temperature inaccuracy issue from the process variation could be resolved.

5.4 Dual Threaded Control Flow

As discussed in the previous section, the purpose of the on-chip heater and the temperature sensor is to maintain a consistent temperature for reliability tests. For reliability experiments, measurement under accurate temperature is critical since it changes the device or interconnect characteristics (e.g., threshold voltage, subthreshold slope, resistance… etc.). However, as shown in Fig. 5.6 (bottom), heater and DUT temperatures are not always stable during the experiment. Especially for the metal heaters whose current direction should be changed periodically to prevent EM in the heater wires, the temperature fluctuates right after switching the current direction. Therefore, for accurate reliability characterization, measurement should be temporarily stopped during the unstable region and resumed once the temperature becomes stable. One way to

implement such a test flow is to separate the heater control loop and the data measurement loop. Fig. 5.6 (top) explains the EM test flow with dual threaded control. The first loop (heater control) runs the heater to raise the DUT temperature. If the temperature is stable (e.g., within ± 1°C from the target temperature for 10 consecutive times), this loop signals the measurement loop, which controls the on-chip data scanning

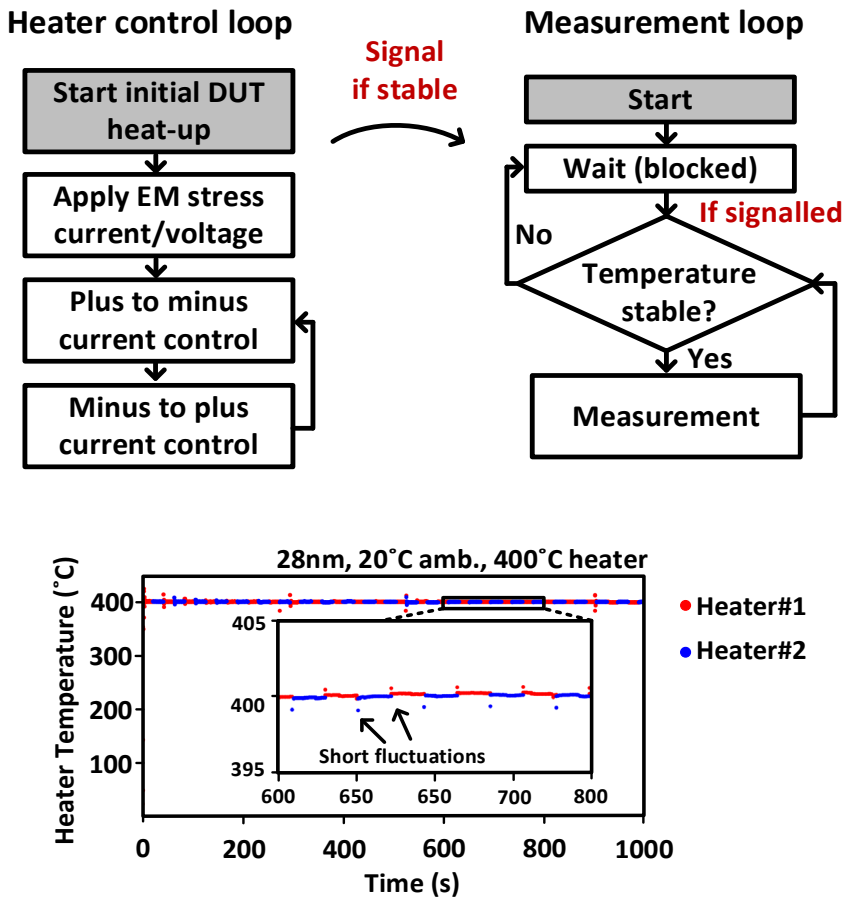


Fig. 5.6. 28nm EM test chip experiment flow with two concurrent controls (top). The heater temperature of with the current direction switches every 20 seconds (bottom)

circuitries and external measurement instruments. One can define a flag that indicates temperature stability. For example, if the temperature is unstable, the flag is set to false so that the measurement loop becomes a wait state and is blocked until the next signal is received from the heater control loop. Using the dual threaded control techniques, the reliability data collection is only done under consistent temperatures, which provides accurate lifetime information and statistics.

5.5 Multiple Heater Control and Multi-Variable Control Issue

If a test structure requires a large heating area and each heater can only cover part of the DUT region, multiple heaters have to be incorporated. The heaters are often placed adjacent to each other to create a uniform stress temperature over the entire DUT area. Fig. 5.7 (a) shows an example of three parallel silicided poly heaters in 65nm technology. The basic idea of operating a multi-heater system is the same as single heater control.

One heater control challenge we faced with the multi-heater system was the thermal interference between the individual heaters. During the measurement, it was found that even with well-calibrated control flows, in some cases, on-chip heaters cannot reach a stable temperature if multiple heaters are closely placed due to the multi-input, multi-output control issue. For example, without any thermal interference, the temperatures of adjacent heaters should depend solely on their own current. However, as illustrated in Fig. 5.7, the temperature of heater #2 is not only decided by current #2 but also by the “side-channel” heat from heater #1, which also depends on current #1. Thus,

if one of the heaters starts to control the adjacent heater's temperature and eventually dominates the entire heating area (Fig. 5.7 (b)), it becomes difficult to achieve a uniform temperature condition.

The main reason for the multi-variable control issue is the inevitable thermal interference between heat sources. Such convoluted interference could be mitigated by lowering the ambient temperature or using a better heat sink, which helps lower the thermal interference.

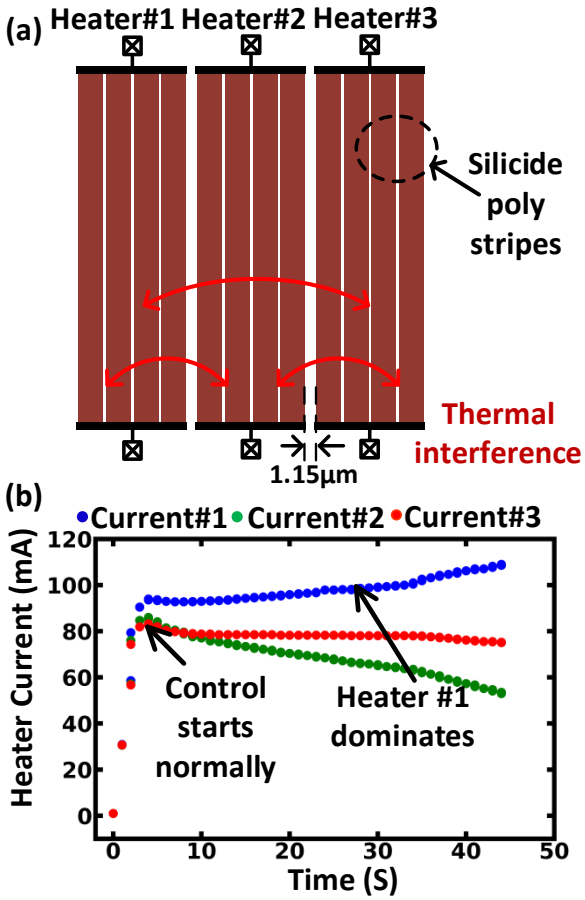


Fig. 5.7. (a) Thermal interferences between heaters can cause thermal runaway issues and prevent stable temperature control. (b) Measurement results show the leftmost heater dominating the temperature control. This issue was resolved by designing a new board with better thermal conductance and lowering the ambient temperature.

5.6 Metal Heater Lifetime and Reliability Concerns

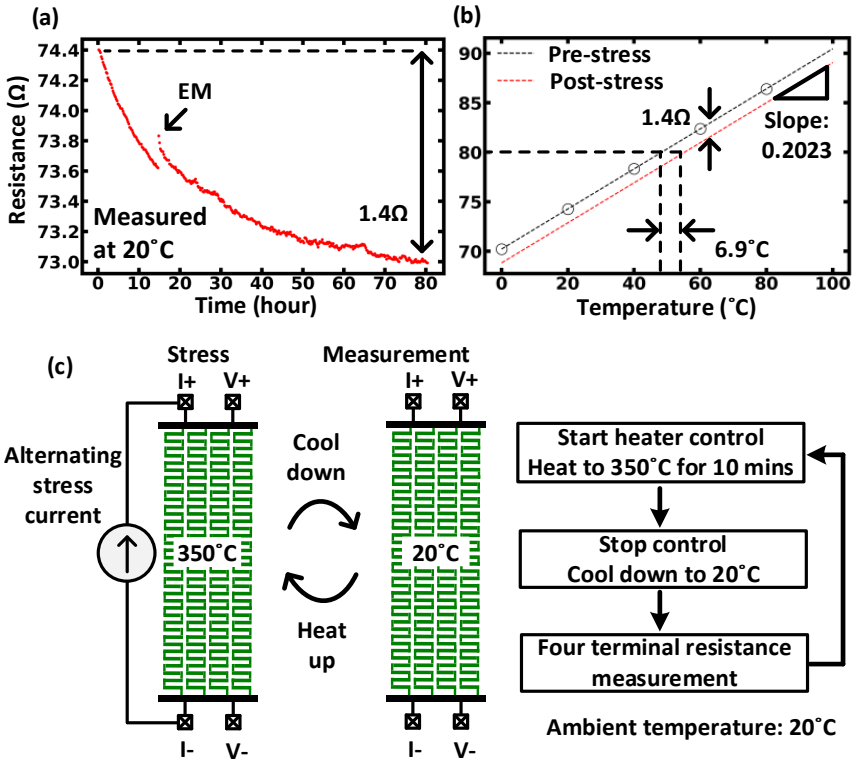


Fig. 5.8. (a) Metal heater resistance's change over its lifetime in a 28nm technology test chip. (b) Impact of heater degradation on temperature measurement accuracy. (c) Heater lifetime experiment.

Due to the narrow snake-shaped wire geometry and the high current density used to reach high temperatures, metal heaters generally face reliability issues. Such heater lifetime issues manifest as an increase or decrease in heater resistance. Fig. 5.8 (c) shows the heater lifetime experiment flow. In this experiment, we operated the heater normally but periodically turned off the heater and re-check its resistance at ambient temperature. The metal heater is stressed by a current higher than 70mA with alternating directions. Fig. 5.8 (a) shows that the metal heater's resistance gradually decreases by 1.4Ω over 80 hours of heater runtime. Also, in the middle of the resistance degradation, an EM-induced

abrupt resistance jump occurs around 15 hours of heater operation. Both heater damages shift the TCR, resulting in inaccurate temperature measurement. For example, as illustrated in Fig. 5.8 (b), a 1.4Ω resistance decrease translates to a 6.9°C temperature overestimation. Thus, after 80 hours of experimentation, the heater temperature is 356.9°C, which is slightly higher than the target. One approach to address the resistance drift issue is recalculating the TCR right before starting the heater control. Since the slope of the TCR generally doesn't change even after the resistance shift, by measuring the resistance at 20°C, TCR could be recalibrated based on the new data point. However, instead of recalibrating the TCR, the best solution is using the DUT temperature-based control with a dedicated temperature sensor, as described in section 5.3. By using such a strategy, heater damages and inaccuracies can be overcome.

Summary

The on-chip heater is useful for raising local die temperature, which is necessary for reliability tests. The introduced negative feedback-based heater current control can increase the DUT temperature in a short amount of time (e.g., 10s seconds). There are several sources of temperature inaccuracy that come from variations in DUTs and the thermal conductivity of dies. For more accurate DUT temperature regulations and reliable data collection, several advanced control methodologies employing temperature sensors and dual-threaded control flow are recommended. Also, multi-variable control issues and the heater's reliability are potential control difficulties.

Chapter 6 Summary

In this thesis, circuit-based EM test chip design and experiment methodologies are proposed. EM is not a trivial reliability concern in modern process technology. High current density in metal routings and high power dissipation in a VLSI system with the scaling down of transistor feature sizes can accelerate the EM-induced BEOL aging behaviors in power supply nets and circuit interconnects. One of the main reliability concerns is a PDN with a high unidirectional current. Since EM failure is significantly faster in DC conditions than AC with alternating current direction, the lifetime of a power grid should be carefully considered, both for the designers and EM aging model developers. In this context, this work presents test chip based power grid EM failure characterization methodologies that allow flexible DUT designs and efficient measurements. Such EM data collected from silicon measurements are also used to calibrate the physics-based EM simulators. Also, this thesis explains the on-chip heater that is widely used in EM reliability testing, including their design/control details and advanced experiment methodologies.

In Chapter 2, basic EM aging mechanisms are introduced to help understand the importance of test chip based reliability data measurement approaches. Several widely used EM models are also explained to show the physical parameters that affect the EM lifetime in intuitive ways. However, simple aging models have limitations on predicting the lifetime of complicated interconnect structures with many redundancies and complex current distributions. Therefore, the physics-based EM simulation concept, which focuses

on stress evolution and void growth in the interconnects, is introduced. Since the silicon measurement can contribute to calibrating the model used for such software, the model developer's approaches to improving the EM EDA tool are briefly explained.

Chapter 3 explains the impact of temperature gradient on a power-grid-like test structure. Metal interconnects in local hotspots on a die (e.g., 10s˚C higher than adjacent regions) have higher atomic diffusivity with faster migration. Such locally accelerated migration generates additional atomic flux divergence at the temperature gradient regions, which can change the EM failure behavior. Given the physical model and problem statement, we compared the EM lifetime and failure locations in a 9x9 metal mesh structure under three different temperature gradient conditions. The test chip's temperature gradients are calibrated by three on-chip heaters to generate (i) uniform temperature, (ii) 20˚C gradient, and (iii) 30˚C gradient conditions. The comparison of the three conditions showed that the temperature gradients in the DUT accelerate the void nucleation, changing the EM failure times and locations.

Chapter 4 discussed the EM experiment on a realistic power grid structure. Using the design flexibility allowed by the test chip and circuit-based design and experimental methodology, four auto PnR-generated power grids with logic cell equivalent resistive loads are implemented. The four grids have different interconnect structures, including cell via counts, power rail width, power rail via redundancy, and rail density. Constant currents and voltages are applied to see the power delivery failure after EM occurs inside the power grid. Thanks to the 1024 voltage tapping capability in various locations in the power nets, the EM failure times and locations are monitored. Also, the metal on-chip

heaters and the temperature sensor allowed experimenting under multiple temperatures. Using the temperature control capability, the grid's lifetime under multiple temperatures and stress currents is presented to roughly estimate the lifetime of the grid. As expected, the strongest grid with wider rail width, rail via counts, and high rail density showed the longest lifetime. Also, high temperature and the current experiments showed accelerated aging. However, the via redundancy of each quasi-cell didn't have a noticeable impact on the failure time. The state-of-the-art physics-based EM simulator validated that the ground nets have earlier void nucleations due to faster material depletion and tensile stress growth.

Finally, in Chapter 5, advanced on-chip heater control methodology is discussed. Even though the heater temperature could be accurately regulated by the introduced negative feedback loop, several factors, such as thermal conductance variation or DUT process variation, make it difficult to maintain a consistent DUT temperature. As a solution for accurate DUT temperature and clean reliability data collection, a dedicated temperature sensor and dual-threaded control flow are proposed. Also, possible difficulties such as multi-variable control issues and heater reliability concerns were introduced.

Bibliography

[1] J.R. Black, “Electromigration failure modes in aluminum metallization for semiconductor devices”, *Proc. of the IEEE*, 1969

[2] V. Sukharev, A. Kteyan, F. Najm, et al., "Experimental Validation of a Novel Methodology for Electromigration Assessment in On-chip Power Grids", *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 2021

[3] A. Kteyan, V. Sukharev, and C.H. Kim, "Novel Methodology for Temperature-Aware Electromigration Assessment in On-chip Power Grid: Simulations and Experimental Validation", *International Reliability Physics Symposium (IRPS)*, 2022

[4] K. Weidge-Zaage, D. Dalleau, X Yu, et al., “Static and dynamic analysis of failure locations and void formation in interconnects due to various migration mechanisms”, *Materials Science in Semiconductor Processing*, 2003

[5] J. Tao, N.W. Cheung, and C. Hu, “Metal electromigration damage healing under bidirectional current stress”, *IEEE Electron Device Letters*, 1993

[6] C. Zhou, X. Wang, R. Fung, et al. “A Circuit based Approach for Characterizing High Frequency Electromigration Effects”, *IEEE Trans. On Device and Materials Reliability (TDMR)*, 2018

[7] C. Zhou, X. Wang, R. Fung, et al., “High Frequency AC Electromigration Lifetime Measurements from a 32nm Test Chip”, *IEEE Symposium on VLSI Technology*, 2015

[8] N. Pande, C. Zhou, MH Lin, et al., “Characterizing Electromigration Effects in a 16nm FinFET Process Using a Circuit based Test Vehicle”, *IEEE International Electron Devices Meeting (IEDM)*, 2019.

[9] N. Pande, C. Zhou, MH Lin, et al, “A 16nm All-digital Hardware Monitor for Evaluating Electromigration effects in Signal Interconnects through Bit-Error-Rate Tracking”, *IEEE Trans. On Device and Materials Reliability (TDMR)*, 2022

[10] N. Pande. C. Zhou, MH Lin, et al., “Electromigration-Induced Bit-Error-Rate Degradation of Interconnect Signal Paths Characterized from a 16nm Test Chip”, *IEEE Symposium on VLSI Technology*, 2021

[11] B. Li, A. Kim, P. McLaughlin, et al., “Electromigration characteristics of power grid like structures”, *IEEE International Reliability Physics Symposium (IRPS)*, 2018

[12] C. Zhou, R. Wong, S. Wen, et al., “Electromigration Effects in Power Grids Characterized Using an On-Chip Test Structure with Poly Heaters and Voltage Tapping Points”, *IEEE Symposium on VLSI Technology*, 2018

[13] C. Zhou, R. Fung, S. Wen, et al., “Electromigration Effects in Power Grids Characterized from a 65nm Test Chip”, *IEEE Trans. on Device and Materials Reliability (TDMR)*, 2019

[14] P. Paternoster, A. Maki, A. Hernandez, et al., “XBOX Series X: A Next-Generation Gaming Console SoC”, *IEEE International Solid-State Circuits (ISSCC)*, 2021

[15] H.V. Nguyen, C. Salm, B. krabbenborg, et al., “Effect of thermal gradients on the electromigration life-time in power electronics”, *IEEE International Reliability Physics Symposium (IRPS)*, 2004

[16] S. Torosyan, A. Kteyan, V. Sukharev, et al., “Novel physics-based tool-prototype for electromigration assessment in commercial-grade power delivery networks”, *Journal of Vacuum Science and Technology B*, 2021.

[17] H. Zahedmanesh, P. Roussel, I. Ciofi, et al., “A pragmatic network-aware paradigm for system-level electromigration predictions at scale”, *IEEE International Reliability Physics Symposium (IRPS)*, 2023

[18] H. Yu, Y. Yi, N. Pande, et al., “On-chip Heater Design and Control Methodology for Reliability Testing Applications Requiring over 300°C Local Temperatures”, *IEEE Trans. Device and Material Reliability (TDMR)*, 2023

[19] O. Kraft, et al., “Quantiative Analysis of Electromigration Damage in Al-based Conductor Lines”, *Journal of Materials Research*, 1997

[20] V. Mishra, et al., “Predicting Electromigration Mortality Under Temperature and Product Lifetime Specifications”, *Design Automation Conference (DAC)*, 2016

[21] Frank L. Wei, et al. “Effects of active atomic sinks and reservoirs on the reliability of CU/low-k interconnects.”, *Journal of Applied Physics*, 2008

[22] M. Hauschildt, et al., “Large-scale statistical analysis of early failures in Cu electromigration, Part I: Dominating mechanisms”, *Journal of Applied Physics*, 2010

[23] I. A. Blech, “Electromigration in thin aluminum films on titanium nitride”, *Journal of Applied Physics*, 1976

[24] J. Lienig, “Electromigration and Its Impact on Physical Design in Future Technologies”, *International Symposium on Physical Design (ISPD)*, 2013

[25] J. Lienig, et al., “Fundamentals of electromigration”, *Springer International Publishing*, 2018

[26] G. Mokry, et al., “High Ampacity Carbon Nanotube Materials”, *Nanomaterials*, 2019

[27] M.A. Korhonen, et al., “Stress evolution due to electromigration in confined metal lines”, *J. Appl. Phys.*, 1993