Full-Chip Electromigration Assessment: Effect of Cross-Layout Temperature and Thermal Stress Distributions

Xin Huang*, Valeriy Sukharev †, Jun-Ho Choy†, Haibao Chen‡, Esteban Tlelo-Cuautle§ and Sheldon X.-D. Tan*
* Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA
† Mentor Graphics Corporation, Fremont, CA 94538, USA
‡ Department of Micro/Nano-electronics, Shanghai Jiao Tong University, Shanghai 200240, China
§ Electronics Department, INAOE, Mexico

***Abstract*—Many prior works have investigated electromigration (EM) on full-chip power grid interconnects, which has become one of major reliability concerns in nanometer VLSI design. However, most of the published results were obtained under the assumption of uniformly distributed temperature and/or residual stress across interconnects. In this paper, we demonstrate the implementation of novel methodology and flow for full-chip EM assessment on the multi-layered power grid networks of a 32nm test-chip and investigate the impacts of the within-die temperature and thermal stress variations on the failure rate. The proposed approach is based on recently developed physics-based EM models and the EM-induced IR-drop degradation criterion that replaces the traditional conservative weakest segment method. The cross-layout temperature distribution caused by power dissipations in devices and by interconnect Joule heating has been characterized and taken into account in the full-chip EM assessment methodology. Results of the simulations performed on the analyzed multi-layered power/ground nets show that traditional assumption of the uniform average temperature leads to inaccurate predictions of the time-to-failure (TTF). Furthermore, the consideration of thermal stress variation results in a retarded EM induced degradation.**

I. Introduction

The continuous shrinking of features sizes in modern ICs has aggregated the EM-induced reliability treats. The power delivery networks are more susceptible to EM effects than the signal lines as they carry large unidirectional currents, and thus can fail in much shorter times due to continuous stress buildup under EM action [1]. Recently, a number of works have been proposed to employ the EM-induced IR-drop degradation analysis as an EM assessment for power grid networks. This analysis considers the inherent redundancy of mesh-structured networks and power-delivery functionality of the grid [1–3], which replaced the traditional conservative weakest segment criterion. Currently employed Blech limit [4] (for the out filtration of immortal segments) and Black's equation [5] (for calculating MTTFs for segments characterized by known current densities and temperatures) are subjects for the hard criticism [6–9]. Cross-die variation of residual stress makes the Blech's "critical product" to be layout dependent variables rather than experimentally determined constants. The widely accepted methodology of calculating the MTTF at use condition is represented by chip operation current density and temperature, while using and determined at the stressed (accelerated) condition, characterized by high current densities and elevated temperatures. However, the interdependency of the Black's activation energy (E_a) and current density exponent (n) on the current density and temperature makes this method rather controversial. The recently proposed new physics-based EM model for void initiation and evolution responsible for a time-dependent resistance degradation of the p/g nets can overcome all the flaws related to the Black-Blech formulation [1, 9].

This work is supported in part by NSF grant under No. CCF-1255899, in part by NSF Grant under No. CCF-1527324, in part by Semiconductor Research Corporation (SRC) grant under No. 2013-TJ-2417.

EM effect on a metal wire is governed by a combination effect of current density, local temperature and residual stress, as shown in [1, 9–11]. The rapid increase in chip power dissipation and power density results in thermal hot spots and temperature gradients on the die. Thermal stress is a major source of the residual stress, which is mainly due to the significant mismatch in the coefficients of thermal expansion (CTE) of the metal and the silicon surrounding it. Therefore in order to accurately estimate the risk of EM induced failure, the full-chip EM assessment methodology should not only account for the current densities but also consider the temperatures and residual stresses in different interconnect segments for different workloads. There have been many prior works focusing on EM analysis techniques under various scenarios. However, unfortunately, most of the published results were obtained under the assumption of uniformly distributed temperature and/or uniformly distributed residual stress across the chip [1–3, 12]. These results will be not accurate enough to characterize the kinetics of void nucleation and void volume evolution responsible for EM-induced degradation and, hence, result in ineffective reliability optimization techniques.

This paper describes the implementation of novel methodology and flow for full-chip EM assessment for multi-layered power grids, considering the impacts of the cross-layout temperature and thermal stress variations on the failure rate. The proposed approach is based on recently developed physics-based EM models and the EM-induced IR-drop degradation criterion that replaced the traditional conservative weakest segment method. We conduct the experiments on a real 32nm test-chip. The main contributions are: (i) Characterization of power dissipation in devices and Joule heating in the interconnects, and the estimation of full chip temperature and thermal stress distribution by compact modeling; (ii) EM assessment of the power grids, which takes into account void volume saturation and considers the within-die temperature and thermal stress variations.

II. Physics-Based Electromigration Modeling

EM effect on metal wires results in metal density depletion or accumulation, which is accompanied by development of the corresponding tensile or compressive hydrostatic stresses at the locations characterized by the atomic flux divergences, for example, at the ends of the metal lines embedded in inter-layer/inter-metal dielectrics (ILD/IMD). The recently proposed physics-based EM models for void initiation and evolution that cause a time-dependent degradation of the segment electrical characteristics allow accurate prediction of interconnect failure rate caused by impacts of current density, temperature and residual stress [1, 9, 13]. From the view of EM-induced degradation, the modern power grid networks consist of a set of interconnect elemental reliability units, also known as interconnect trees [14], representing continuously connected, highly conductive metal lines within one layer of metallization, terminated by diffusion barriers. The absence of blocking

boundaries at one or both ends of the branches allows atoms to transfer between connected limbs in a tree preventing them from accumulation/depletion, and hence, eliminates related stress buildup at the branch connections. Thus the traditional immortality assessment and the time-to-failure (TTF) calculation methods, which are based on a single interconnect line confined by IMD/ILD dielectric, is not applicable for the power grid analysis. Instead, the tree-based analysis deals with the grid structure and accounts for the stress redistribution between the connected branches in a tree should be performed, [1, 14].

The EM process can be characterized as a sequence of two events: *void nucleation* and a sub-sequencing *void growth*:

- **Void nucleation phase:** Void nucleation kinetics in the interconnect segment can be derived from the solution of kinetics equation describing the time evolution of stress in the interconnect segment [10, 15–17]. Void nucleation time (t_{nuc}) is determined as an instant in time when stress at the cathode end of the line, characterized by the biggest tensile stress, reaches the critical stress (σ_{crit}) and can be calculated as [1]:

$$t_{\text{nuc}} \approx \tau^* e^{\frac{E_V}{kT}} e^{-\frac{f\Omega}{kT}(\sigma_T+\sigma_{\text{EM}})} ln\left\{\frac{\sigma_{\text{EM}}}{\sigma_{\text{EM}}+\sigma_T-\sigma_{\text{crit}}}\right\} \quad (1)$$

  where $\tau^* = \frac{L^2}{2D_0} e^{\frac{E_D}{kT}} \frac{kT}{\Omega B}$, the steady state EM stress at cathode end of a wire σ_{EM} is a function of current density j and wire length l, and $\sigma_{\text{EM}} = \frac{eZ\rho jL}{2\Omega}$ for a 1D confined line. Here, E_{V} and E_{D} are the activation energy of vacancy formation and diffusion, f is the ratio of volumes occupied by a vacancy and a lattice atom, σ_T is the thermal stress developed in the metal line confined in the ILD/IMD environment during cooling from the zero stress temperature T_{zs} down to the temperature of use condition, B is the bulk modulus [1, 9]. For an interconnect tree, the hydrostatic stress (including EM stress and residual stress) at all nodes of a tree should be obtained based on Eq. (2) and Eq. (3), under a void-less steady state assumption, [1]:

$$\sigma_i^c - \sigma_j^a = \triangle\sigma_{ij} = -\frac{eZ\rho(j_{ij}L_{ij})}{\Omega} \quad (2)$$

$$\sum_{i=1}^{k}\left(\sigma_i - \left[\sigma_T + \frac{eZ\rho(j_{ij}L_{ij})}{2\Omega}\right]\right)L_{ij} = 0 \quad (3)$$

  where σ_i^c and σ_j^a are the hydrostatic stresses at the cathode and anode ends of the ij-branch.
- **Void growth phase:** Critical void formed at t_{nuc} grows its volume above critical one at $t > t_{\text{nuc}}$. The kinetics of the void volume evolution governs the evolution of the segment electrical resistance. In the growth phase, the resistance of a void containing branch increases over time. As a result, the p/g network becomes a time-varying network and its IR drops will keep changing over time. The kinetics of line resistance change can be approximately described as [1, 13]:

$$\Delta R(t) = \vartheta(t - t_{\text{nuc}})\left[\frac{\rho_{\text{Ta}}}{h_{\text{Ta}}(2H+W)} - \frac{\rho_{\text{Cu}}}{HW}\right] \quad (4)$$

  where $\vartheta = \frac{DeZ\rho j}{kT}$ is the void edge drift velocity, $D = D_0 e^{-\frac{E_D}{kT}}$ is the atom diffusivity. Here ρ_{Ta} and ρ_{Cu} are the resistivities of the barrier material (T_a/T_aN) and copper, W is the line width, H is the copper thickness, and h_{Ta} is the barrier layer thickness.

It should be noted that the void ***saturated volume*** develops when growing void consumes all volumetric deformation generated by thermal stress and by redistribution of atoms removed from the space occupied by growing void. Once the void volume becomes saturated, the further resistance change is stopped until the current density will be changed due to other voids developed somewhere inside same interconnect tree or in other trees. The saturated void volume can be calculated as $V_{\text{SS}} = V_{\text{line}}(\sigma_T + \sigma_{\text{EM}})/\Omega$, where V_{line} is the volume of the metal line [18].

III. Cross-layout temperature and thermal sterss characterization

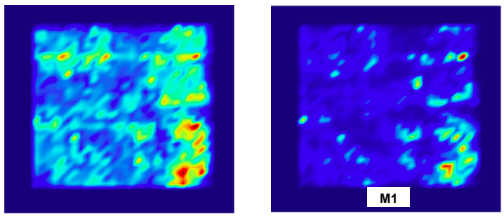
This section describes the estimation of within-die temperature and thermal stress distribution. We demonstrate the characterization of power dissipation in the devices and the Joule heating in interconnects, then, we detail the temperature simulation methodology, which builds a thermal netlist with the effective thermal properties. Further, the thermal stress variation across the layout will be obtained based on the temperature distribution. A 32nm test-chip is used as the case example to present the flow. Standard-cells are used in this design along with 7 metal (copper) layers. The dimension of the layout is $184\mu m \times 184\mu m$, and there are 16 layers in total (the Si-layer is divided into a thin Si-device layer that includes power dissipation and a thick Si-substrate layer for thermal analysis purpose).

A. Full-chip power characterization

Power estimation is a challenging task since the result is highly dependent on the workload or typical usage of a chip. We consider both the power dissipation in the devices and Joule heating in interconnects. The chip power consists of static and dynamic components. Eq. (5) is the power model for each individual primitive block, where a block is considered to be a primitive if it cannot be further decomposed into smaller blocks based on an user settings. The dynamic power comes from the gate switching, in which the block dissipates power by charging the load capacitances of wires and gates and dissipates power during a very short period of time when a conduction path exists between the power and ground voltage connections. Thus, more active block is characterized by the higher dynamic power. In contrast, the static power is due to static current, including the leakage current and presents regardless of a block's activity level.

$$P_{\text{block}} = P_{\text{static}} + \alpha_{\text{switching}} \times P_{\text{dynamic}} \quad (5)$$

Here, P_{static} and P_{dynamic} are the static and dynamic powers of a block, $\alpha_{\text{switching}}$ is the switching estimate for signals that describes the switching activity of the block.



(a) (b)

Fig. 1. Power consumption in device layer (0-6.91mW), (a), and Joule in M1 layer (0-0.06mW), (b).

For time varying currents, the time scale is on the order of picosecond, which is too fine for the thermal time scale. As a result, power averaging is further applied to obtain the power consumption in the device layer. The current flowing through interconnects generates Joule heating. Similar to Eq. (5), we estimate the Joule heating in a wire by evaluating both static and dynamic components of generated heat. Fig. 1 selectively shows the power map for the device layer and the Joule heating in M1 layer.

B. Thermal simulation methodology

Since the temperature affects atom diffusivity, the EM assessment requires accurate local temperature estimation at each interconnect layer in order to adequately account for the temperature-sensitive void nucleation and growth kinetics. In this section, we describe a thermal analysis flow that efficiently estimates the cross-die temperature variation by employing a compact thermal model that represents a die as arrays of cuboidal thermal cells with effective local thermal properties. The methodology includes three steps: (i) extract effective thermal properties of a thermal cell in a layer, (ii) generate thermal netlist of the whole chip, and (iii) calculate temperature at each thermal node by a circuit solver. All considered composite layers are divided into a set of thermal cells defined by a layer thickness and a square window of size l which is chosen based on the simulation accuracy: the finer partitioning provides more accurate results at the expense of the run time. Each cell contains 6 thermal resistors representing heat propagation in the lateral and vertical directions; a thermal capacitance can be included for transient thermal analysis.

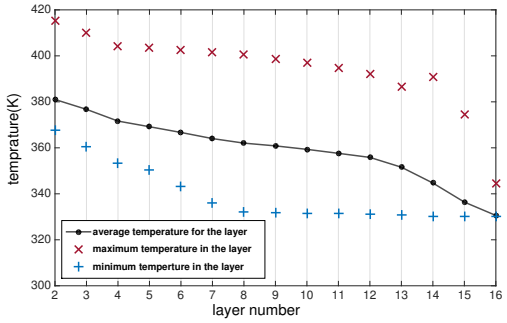


Fig. 2. Temperature variation across different layers

The effective thermal conductivities are functions of metal density and routing direction of wires in each metal layer based on the theory of effective thermal properties of anisotropic composite materials [19]. Based on the standard procedure [20, 21], with the extracted thermal resistances/capacitances, estimated power sources, as well as the thermal boundary conditions, the chip can be represented as a thermal netlist, in which the nodal temperatures correspond to the nodal voltages and the powers corresponds to the currents. The electric circuit solver can then obtain temperature for each thermal node. In the thermal simulation, performed in the analyzed design, a window size $l = 5\mu$m is chosen for computational efficiency while keeping reasonable temperature resolution. The top surface of the die is kept at T=330K, while all other sides are insulated. As shown in Fig. 2, temperature varies across different layers. The temperature distribution in M1 interconnects is demonstrated in Fig. 3.

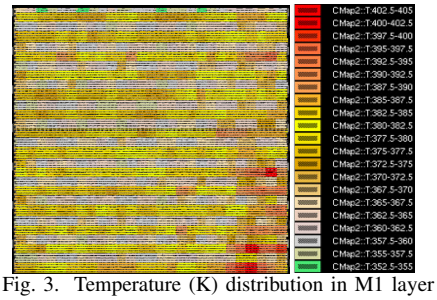


Fig. 3. Temperature (K) distribution in M1 layer

C. Thermal stress estimation

Since the residual stress existing in each metal line before the electrical load is applied is responsible for both void nucleation and growth, its distribution should be considered for accurate chip-scale EM assessment. This work focuses on the thermal stress, which is a major source of the residual stress. In the case of metal line embedded into the rigid confinement, which, in the analyzed case of the on-chip interconnect, is comprised by refractive metal and dielectric diffusion barriers and ILD/IMD dielectrics, the initial stress is generated during the system cooling from the stress free annealing temperature T_{ZS} down to the use temperature T. A primary source of this thermal stress is the difference in the coefficients of thermal expansion (CTE) of the metal α_{M} and confinement α_{conf}, which is mainly determined by the silicon substrate:

$$\sigma_T = B(\alpha_{\text{M}} - \alpha_{\text{conf}})(T_{\text{ZS}} - T) \tag{6}$$

The thermal stress distribution in the M1 layer is demonstrated in Fig. 4. Since a lower thermal stress corresponds to a higher layout temperature due to the smaller temperature gap $\Delta T = T_{\text{ZS}} - T$, the regions with the highest T, Fig. 3, are characterized by the smallest σ_T, Fig. 4. Note that more accurate thermal stress estimation should consider the elastic portion of the stress generated by interaction with confinement. This is reserved for the future analysis. Nevertheless, the calculated inelastic fraction of the thermal stress demonstrates well the trend caused by the thermal stress variation and its relation to the temperature distribution.

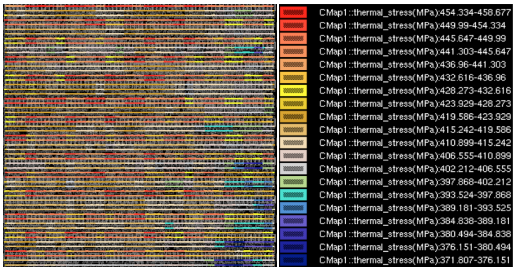


Fig. 4. Thermal stress (MPa) distribution in M1 layer

IV. Experimental Results and Discussion

The proposed full-chip thermal variation aware-EM assessment method is implemented by C++ code on a 2.4GHz Linux server and tested on the 32nm standard-cell IC design. We use Calibre tools for layout extraction and circuit analysis. The EM analysis results can be mapped into the BEOL layout and displayed in Calibre RVE tool. The power net and ground net are symmetrical and independent on each other since the IP blocks are modeled as effective DC current sources. In this work, the thermal analysis targets the whole chip and the EM analysis focuses on the power net. The EM assessment is achieved through following steps: (i) IP block modeling and power grid network extraction, (ii) cross-layout temperature and thermal stress characterization, and (iii) full-chip EM analysis considering temperature and thermal stress distributions, where the simulation steps are similar to those in [1], but saturated void volume effect is accounted and the thermal stress and temperature are no longer uniformly distributed. We assume the chip fails when the maximum IR drop exceeds $10\%V_{\text{DD}}$. Parameters used in EM model are listed in Table I.

TABLE I
Parameters used in EM model

Parameter	Value	Parameter	Value
E_V	0.73eV	Z	10
E_D	0.53eV	σ_{crit}	500MPa
f	0.6	T_{ZS}	623K
B	1×10^{11}Pa	D_0	7.56×10^{-5}m^2/s

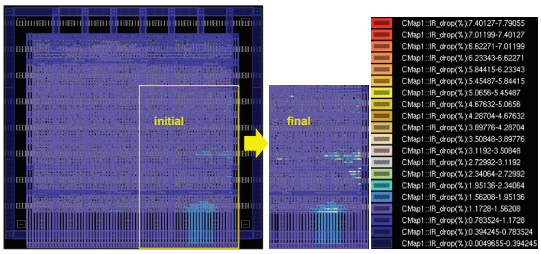
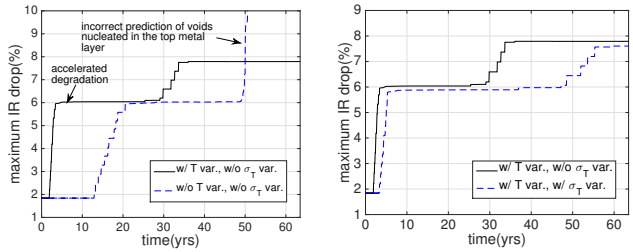
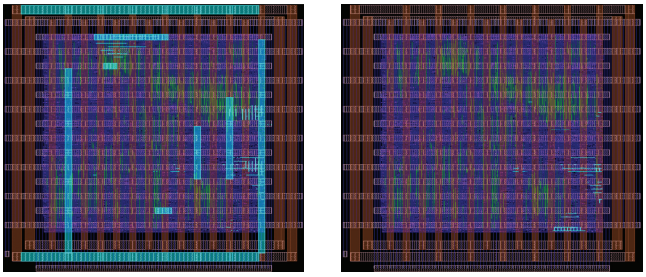


Fig. 5. EM-induced IR drops change in the power net



(a) (b)

Fig. 6. Comparison of predicted maximum IR drop increase between w/o and w/ temperature variation conditions, where uniformly distributed thermal stress is considered in both cases, (a) and comparison of the IR drop evolution calculated w/o and w/ thermal variation assessments, where temperature variation is considered in both cases, (b).



(a) (b)

Fig. 7. The branches with voids (blue color) at the time when simulation stops in two cases: (a) uniform average temperature, and (b) with temperature variation. Uniform thermal stress distribution is considered in both cases.

Fig. 5 shows the EM-induced IR drop increase in the power net. To evaluate the impact of within-die temperature variation, we performed EM assessment under assumption of uniform temperature (T=367.4K, average temperature across the chip) and with the calculated non-uniform temperature distribution. Uniform average thermal stress σ_T=434.55MPa is assumed in both cases. The evolution of IR drop degradation and the locations of branches with voids when the simulation stops, are shown in Fig. 6(a) and Fig. 7 respectively.

We have several interesting observations: First, the accelerated voltage degradation caused by high local temperatures at lower layers is observed in the case of non-uniform temperature distribution during the earlier time. Second, as time goes by, the drastic IR drop increase is found under the average temperature assumption. At this time these two assumptions result in totally different analysis results (the average temperature case causes a false-alarm chip failure). This is mainly due to the incorrect prediction of voids nucleated in the top layers where the actual local temperature is much lower than the average value, which can be viewed by comparing Fig. 7(a) and Fig. 7(b).

Next, we have investigated the impact of thermal stress on EM effects by comparing the IR drop evolution in two cases characterized by non-uniform and uniform (σ_T=434.55MPa) thermal stress distributions. Fig. 6(b) shows the IR drop evolution taking place in these two cases. Contrary to the impacts of spatial temperature variation, the retarded voltage degradation is observed when the thermal stress variation is taken into account. The failure prefers to happen at places where both the hydrostatic stress and the local temperature are high. Since a higher temperature corresponds to a lower thermal stress, when spatial thermal stress variation is considered, the branches with the high temperature have smaller initial stresses compared with the average value, thus a longer evolution time is needed for stress to reach the critical value required for void nucleation.

V. Conclusion

In this work, we have proposed and implemented a new EM assessment method for the multi-layered power grid networks of a 32nm test-chip, based on recently proposed physics-based EM models and the EM-induced IR drop degradation criterion. The proposed method characterizes the within-die temperature variation caused by power dissipation in devices and by Joule heating in interconnects, and investigates its impacts on failure rate of power grid networks. It also accounts the across chip thermal stress variation. The experimental results show that the EM induced degradation is a combined effect of temperature and hydrostatic stress. Employing traditional uniform average temperature assumption leads to inaccurate predictions of the TTF. It also reveals that the consideration of thermal stress variation results in a retarded voltage degradation.

References

[1] X. Huang, T. Yu, V. Sukharev, and S. X.-D. Tan, "Physics-Based Electromigration Assessment for Power Grid Networks," in *Proc. Design Automation Conf. (DAC)*, pp. 1–6, 2014.

[2] S. Chatterjee, M. B. Fawaz, and F. N. Najm, "Redundancy-Aware Electromigration Checking for Mesh Power Grids," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2013.

[3] V. Mishra and S. S. Sapatnekar, "The Impact of Electromigration in Copper Interconnects on Power Grid Integrity," in *Design Automation Conference (DAC), 2013 50th ACM/EDAC/IEEE*, 2013.

[4] I. A. Blech, "Electromigration in thin aluminum films on titanium nitride," *Journal of Applied Physics*, vol. 47, no. 4, pp. 1203–1208, 1976.

[5] J. R. Black, "Physics of Electromigration," in *12th International Reliability Physics Symposium*, pp. 142–149, IEEE, 1974.

[6] M. Hauschildt, C. Hennesthal, G. Talut, O. Aubel, M. Gall, K. B. Yeap, and E. Zschech, "Electromigration Early Failure Void Nucleation and Growth Phenomena in Cu and Cu(Mn) Interconnects," in *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pp. 2C.1.1–2C.1.6, IEEE, 2013.

[7] M. Ohring, *Reliability and Failure of Electronic Materials and Devices - Milton Ohring - Google Books*. San Diego: Academic Press, 1998.

[8] J. R. Lloyd, "New models for interconnect failure in advanced IC technology," *Physical and Failure Analysis of Integrated Circuits, 2008. IPFA 2008. 15th International Symposium on the*, pp. 1–7, 2008.

[9] V. Sukharev, "Beyond Black's Equation Full-Chip EM/SM Assessment in 3D IC Stack," *Microelectronic Engineering*, vol. 120, pp. 99–105, May 2014.

[10] V. Sukharev, "Physically based simulation of electromigration-induced degradation mechanisms in dual-inlaid copper interconnects," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, pp. 1326–1335, Sept. 2005.

[11] M. Pathak, J. S. Pak, D. Z. Pan, and S. K. Lim, "Electromigration Modeling and Full-Chip Reliability Analysis for BEOL Interconnect in TSV-based 3D ICs," in *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pp. 555–562, IEEE, 2011.

[12] G. Posser, V. Mishra, P. Jain, R. Reis, and S. S. Sapatnekar, "A Systematic Approach for Analyzing and Optimizing Cell-Internal Signa Electromigration," in *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pp. 1–6, IEEE, 2014.

[13] V. Sukharev, X. Huang, H. Chen, and S. X.-D. Tan, "IR-Drop Based Electromigration Assessment: Parametric Eailure Chip-Scale Analysis," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2014.

[14] S. P. Hau-Riege and C. V. Thompson, "Experimental Characterization and Modeling of the Reliability of Interconnect Trees," *Journal of Applied Physics*, vol. 89, pp. 601–609, January 2001.

[15] M. A. Korhonen, P. Borgesen, K. N. Tu, and C. Y. Li, "Stress Evolution Due to Electromigration in Confined Metal Lines," *Journal of Applied Physics*, vol. 73, no. 8, pp. 3790–3799, 1993.

[16] J. J. Clement, "Reliability Analysis for Encapsulated Interconnect Lines under DC and Pulsed DC Current Using A Continuum Electromigration Transport Model," *Journal of Applied Physics*, vol. 82, no. 12, pp. 5991–6000, 1997.

[17] M. E. Sarychev, Y. V. Zhitnikov, L. Borucki, C.-L. Liu, and T. M. Makhviladze, "General model for mechanical stress evolution during electromigration," *Journal of Applied Physics*, vol. 86, no. 6, pp. 3068–3075, 1999.

[18] Z. Suo, *Reliability of Interconnect Structures*, vol. 8 of *Comprehensive Structural Integrity*. Amsterdam: Elsevier, 2003.

[19] R. M. Christensen, *Mechanics of Composite Materials - Richard M. Christensen - Google Books*. New York: Dover Publication, 2005.

[20] W. Huang, M. R. Stan, and K. Skadron, "Parameterized physical compact thermal modeling," *IEEE Trans. on Components and Package Technologies*, vol. 28, no. 4, pp. 615–622, 2005.

[21] A. Sridhar, A. Vincenzi, M. Ruggiero, T. Brunschwiler, and D. A. Alonso, "3D-ICE: Fast Compact Transient Thermal Modeling for 3D-ICs with Inter-Tier Liquid Cooling," in *IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pp. 463–470, 2010.