Coupled Electro-Thermal-Mechanical Modeling of DRAM Cell Reliability Under Aggressive Scaling and Joule Heating

MSE404MAC - Arjun Kulkarni

Department of Materials Science and Engineering

University of Illinois at Urbana-Champaign

December 2025

Abstract

The research creates a finite element modeling system which uses continuum-scale methods to forecast DRAM system reliability when operating at small feature sizes (below 20 nm) and high current densities that produce heat through Joule effect. The method combines electrical heating with heat diffusion and thermoelastic stress analysis to determine temperature increases and thermal stress effects on wordlines and bitlines and their neighboring capacitor dielectric materials. The present reliability assessment techniques evaluate thermal and mechanical failure mechanisms independently by using empirical acceleration factors which fail to predict failures in sub-15 nm nodes because these nodes experience self-heating and coupling effects. The model unites Joule heating from current crowding with thermal diffusion and thermoelastic stress through a single system which uses properties from literature. The continuum approximation works well for features which exceed 5 nm in size and Fourier heat conduction with linear elasticity produces first-order macroscale results. This model enables a priori screening of DRAM interconnect designs for high-performance computing workloads without requiring fabrication. The main risks stem from using basic 2D geometric models and assuming materials have uniform properties and ignoring unique interface characteristics. The simulation process takes about 30

seconds to complete each parameter set which runs on an Intel i7-class laptop using Python/NumPy. The process of running parametric sweeps and creating visualizations needs between 2 to 3 hours of time. The definition of mid-term success requires scientists to obtain matched temperature and stress patterns which show a 10–50% improvement when the system dimensions grow larger. The researchers need to discover particular feature dimensions and current strength combinations which will generate stress values that exceed the SiO_2 dielectric material's fracture point of 2 GPa.

1. Introduction

1.1 DRAM technology and scaling

Modern computing systems depend on Dynamic Random Access Memory (DRAM) as their main memory technology because it enables both fast operation of computing platforms and AI training accelerators. The basic storage element of DRAM consists of a single transistor combined with one capacitor which forms the 1T1C structure. The metallic wordlines (WL) and bitlines (BL) allow current to pass through during read and write operations which produces high current densities that affect nanoscale interconnect cross-sections because of current modern scaling requirements.

The reduction of features for smaller sizes results in higher areal density but produces major reliability issues. Three coupled physical effects become more severe as dimensions shrink. The process of Joule heating in highly scaled interconnects produces more intense heating because the smaller cross-sectional area leads to higher electrical resistance which produces more heat at specific points that increase in proportion to $q = J^2 / \sigma$. The present heat distribution pattern produces additional hotspots which emerge at both geometric boundaries and material interfaces to create a more intense effect. The temperature increase from the process creates mechanical reliability issues because different materials in DRAM

stacks expand at different rates when heated. The expansion of tungsten interconnects exceeds the expansion of surrounding silicon dioxide dielectrics when temperature increases which produces substantial stresses between the materials and within their constrained spaces. The process of electromigration becomes faster because of high current densities which push atoms through metal lines to create voids that eventually cause open-circuit failure. The degradation process which affects interconnects follows Black's equation which shows that current density and temperature determine the lifespan of interconnects.

1.2 Motivation and objectives

The assessment of industrial reliability depends on empirical models which predict lifetime acceleration based on fabrication data (e.g., Black's equation). The strategy becomes less effective when devices operate below 15 nm because self-heating effects become more pronounced and device stack information remains proprietary which prevents TCAD access and multi-physics effects become necessary.

The research develops an open system which models multiple physical phenomena across the entire system size to enable fast virtual testing of DRAM designs before manufacturing. The main concept shows that temperature increase and thermal stress development must be studied together because severe surface roughness creates hotspots from Joule heating which strengthen stress points at material boundaries.

The research aims to achieve four specific goals which include developing mathematical equations that link Joule heating to heat diffusion and thermoelastic stress behavior. The research will use Python with NumPy and SciPy libraries to develop a numerical solution which can handle finite-element or structured-grid calculations. The research will measure how temperature peaks and stress peaks from von

Mises stress respond to changes in feature size between 10 and 20 nanometers and current density between 10^10 and 10^11 A/m^2. The research will evaluate calculated stress values against published failure limits to identify the essential parameters which determine the critical operating conditions.

2. Methods

2.1 Governing equations

2.1.1 Joule Heating

Volumetric heat generation in a conductor is computed from the local current density and electrical conductivity. The heat generation rate is given by

$$q = J^2 / \sigma$$

where q is the volumetric heat generation rate (W/m^3), J is the current density (A/m^2), and σ is the electrical conductivity (S/m). In this work, tungsten is used as the representative interconnect material for DRAM wordlines and bitlines due to its widespread use and high current-carrying capability.

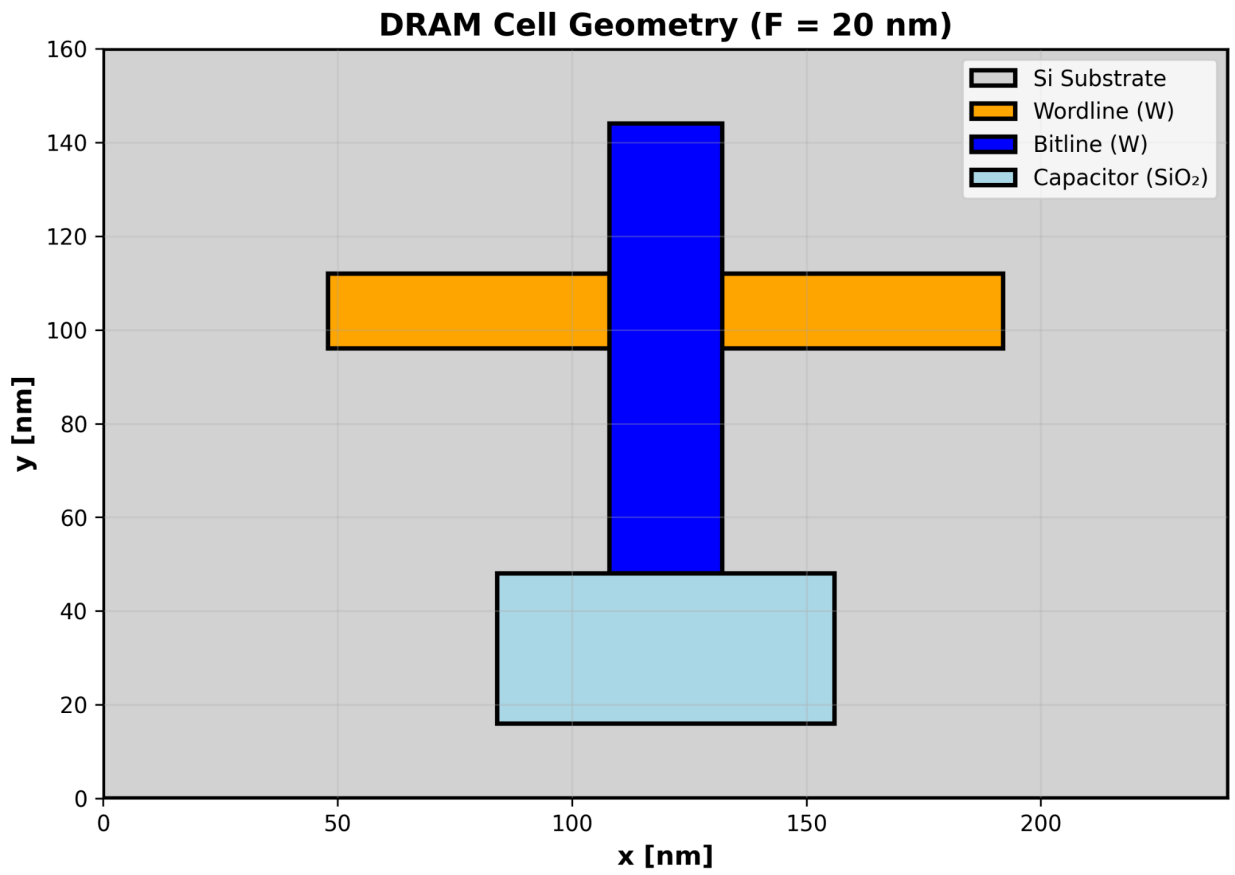

Figure 1: DRAM Cell Geometry

2.1.2 Heat diffusion (steady state)

The heat conduction equation needs to be solved to obtain the steady-state temperature distribution which includes the thermal conductivity (k) and temperature field (T). The thermal conductivity values for tungsten and silicon and silicon dioxide materials have been defined to achieve precise heat distribution modeling in the DRAM stack. The computational domain receives Dirichlet boundary conditions which set the outer boundaries to 300 K fixed

temperature to represent heat exchange with a big heat-sink environment and produce a stable steady-state solution.

2.1.3 Thermoelastic stress

Thermal expansion induces a thermal strain given by $\varepsilon_{th} = \alpha \Delta T$, where α is the coefficient of thermal expansion and $\Delta T = T - T_{ref}$ is the temperature rise relative to a reference temperature. Under isotropic linear elasticity and small-strain assumptions, thermal stress in constrained regions can be approximated using a plane-strain formulation as $\sigma_{thermal} \approx -(E / (1 - \nu)) \alpha \Delta T$, where E is Young's modulus and ν is Poisson's ratio. To assess mechanical reliability, a von Mises equivalent stress is computed from the in-plane stress components and used as a scalar indicator of failure risk within the DRAM stack.

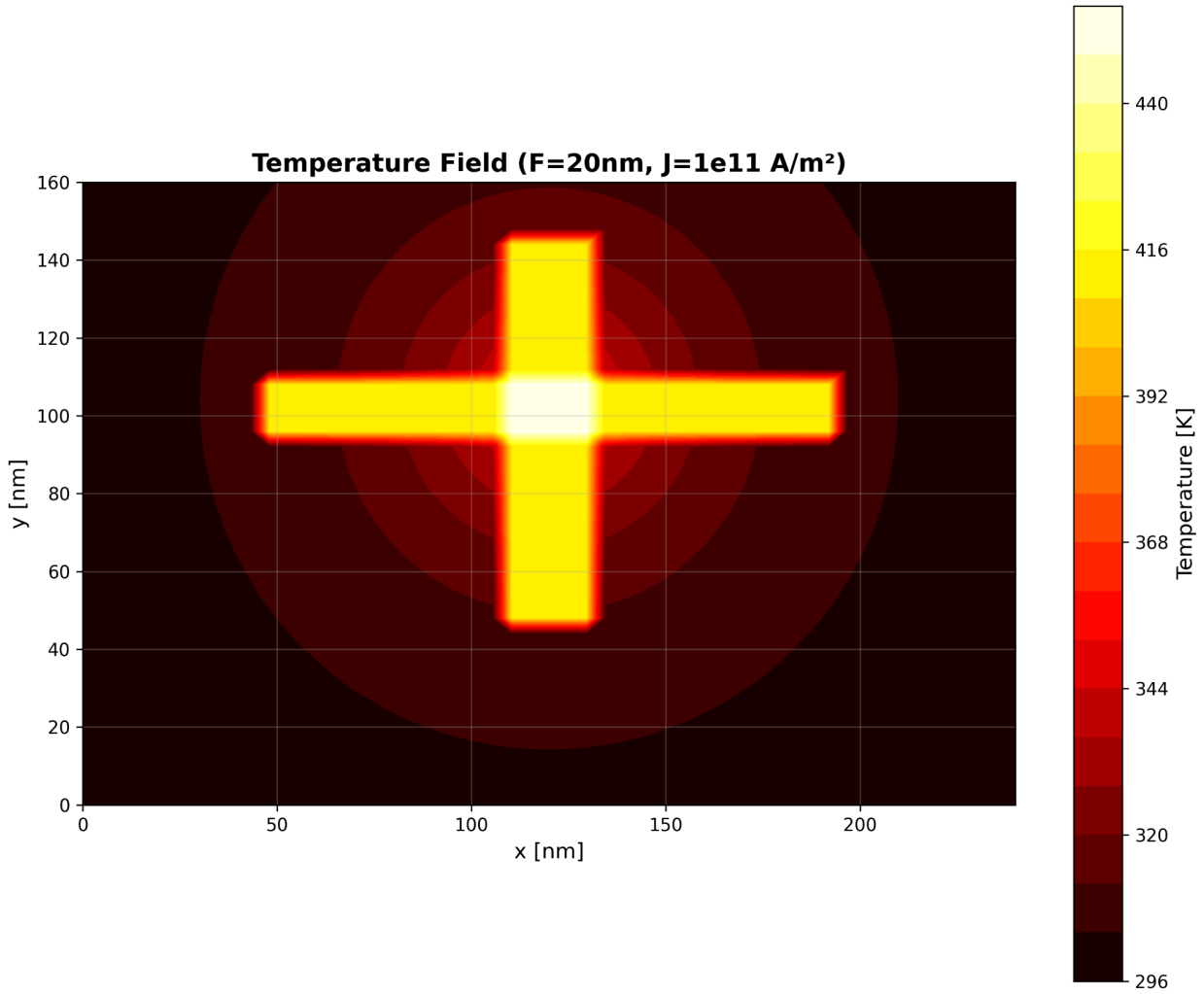

Figure 2: Temperature Field

2.2 Geometry and material properties

2.2.1 DRAM cell cross-section geometry

A simplified two-dimensional cross-section is used to capture the dominant heat generation and interfacial stress behavior in the DRAM cell. The geometry consists of a wordline modeled as a

horizontal tungsten bar with a width of approximately 1F and a height of about 1.5F, and a bitline modeled as a vertical tungsten bar with a width of approximately 1F and a height of about 6F. A capacitor dielectric region is represented as a silicon dioxide (SiO_2) block with dimensions of roughly 2F × 2F, embedded within a silicon substrate and background domain measuring approximately 12F × 8F. Three technology nodes are considered in the simulations, corresponding to feature sizes of F = 20 nm, 15 nm, and 10 nm.

2.2.2 Material properties (300 K)

Representative properties are assigned from standard references on thermophysical properties and thin-film mechanics. The model uses thermal conductivity, density, specific heat (for completeness), elastic constants, CTE, and electrical conductivity for tungsten interconnects.

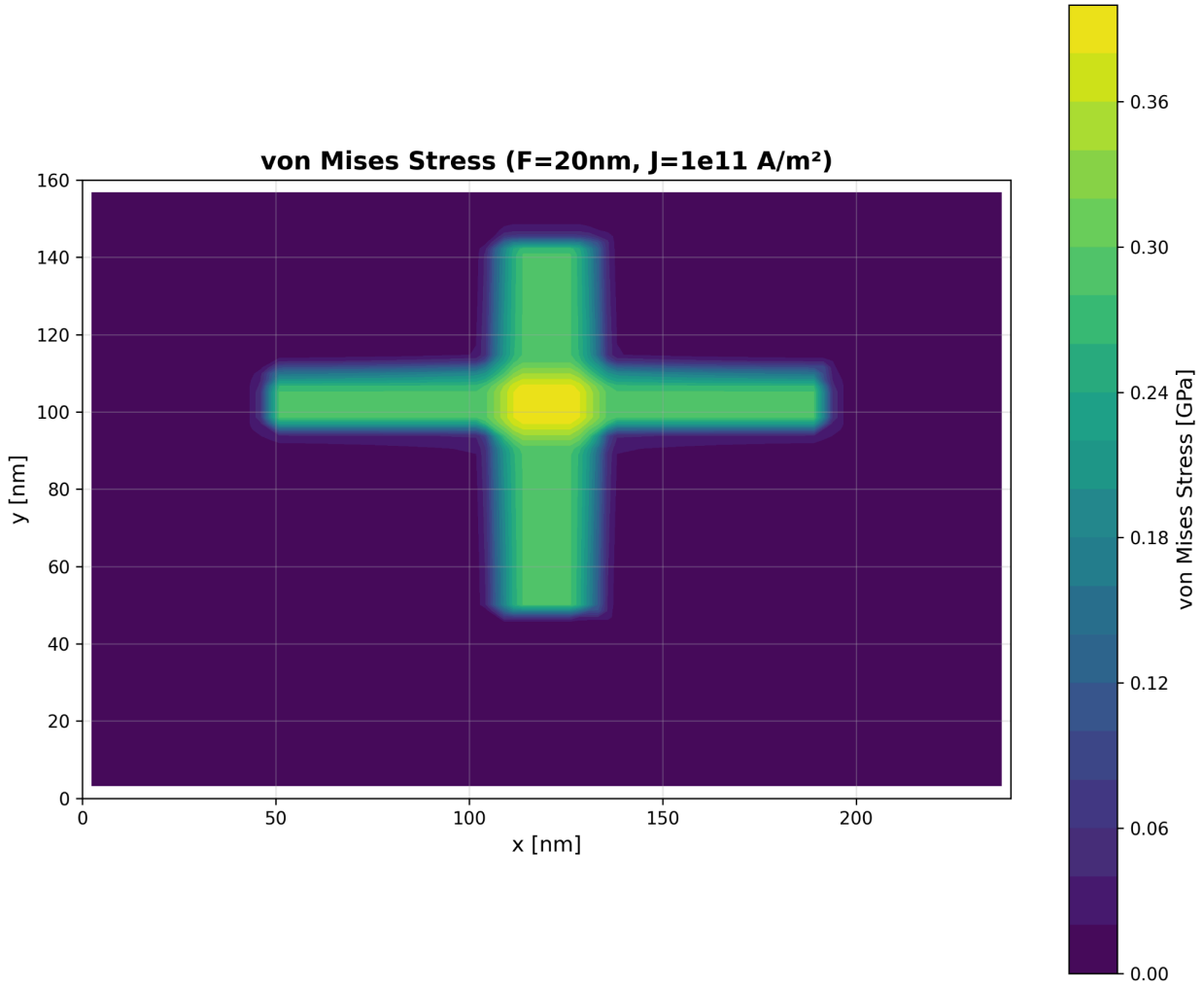

Figure 3: Von Mises Stress

2.3 Numerical implementation

The simulation uses a 50 × 40 two-dimensional quadrilateral grid which produces 2000 elements and 2100 nodes. The analysis requires 2–3 elements which focus on the smallest geometric features because this number provides sufficient detail to track temperature and stress changes. The steady-state heat conduction equation is discretized on the structured grid using a standard

finite-difference/finite-volume stencil, producing a sparse linear system of the form KT = F. This system is solved using sparse linear algebra routines implemented in SciPy. After computing the temperature field, element-level temperature rises (ΔT) are calculated relative to the 300 K reference temperature. Thermoelastic stress is then evaluated using the plane-strain constrained expansion approximation and converted to a von Mises equivalent stress. The process of obtaining nodal values requires stress averaging between elements which share boundaries.

2.4 Parametric study

The simulation performs a complete factorial parameter scan which examines both feature size and current density at three different values for each parameter. The simulation uses three different feature sizes which are F = 20 nm and F = 15 nm and F = 10 nm while it tests three different current densities which are J = 1×10^10 A/m^2 and J = 5×10^10 A/m^2 and J = 1×10^11 A/m^2. The simulation process produces nine different results. The simulations complete their runs within 30 seconds which allows users to quickly study how system dimensions affect operational performance and energy consumption.

3. Results and Discussion

The simulation performs a complete factorial parameter scan which examines both feature size and current density at three different values for each parameter. The simulation uses three different feature sizes which are F = 20 nm and F = 15 nm and F = 10 nm while it tests three different current densities which are J = 1×10^10 A/m^2 and J = 5×10^10 A/m^2 and J =

1×10^11 A/m^2. The simulation process produces nine different results. The simulations complete their runs within 30 seconds which allows users to quickly study how system dimensions affect operational performance and energy consumption.

3.1 Temperature distribution and hotspots

The simulation results show that the highest temperature point appears at the wordline-bitline intersection because current crowding effects occur in this area. The model indicates that the 20 nm F structure will achieve its highest temperature at 380 K when the current density reaches 1×10^11 A/m^2. The peak temperature reaches 380 K which represents an 80 K increase from the surrounding environment. The silicon dioxide region remains cooler than the hottest metal intersection due to its relatively low thermal conductivity and its separation from the primary heat-generation zones. The temperature differences in the system exist mainly at material interfaces and near the point where different materials meet which requires proper mesh density for accurate results. The research supports the idea that thermal gradient differences between materials determine reliability risks more than their average temperatures do.

3.2 Thermomechanical stress localization

The difference in thermal expansion between tungsten and SiO_2 creates high stress concentrations which occur in specific areas near their interface boundaries. The model predicts peak von Mises stress in the SiO_2 dielectric adjacent to heated tungsten features. For F = 20 nm and J = 10^{11} A/m², predicted maximum stress is on the order of ~1.2 GPa. The system experiences tensile stress in its dielectric material because tungsten expands more than SiO_2

when ΔT remains constant which results in compressive stress on the metal material. The model functions as a basic failure detection system which verifies stress levels against predetermined fracture strength intervals which apply to brittle dielectric materials like SiO_2 that break at 1–2 GPa when defects appear and environmental factors impact the material. The predicted stress values under high-current conditions reach the highest values in this range which suggests a possible area where cracking could occur.

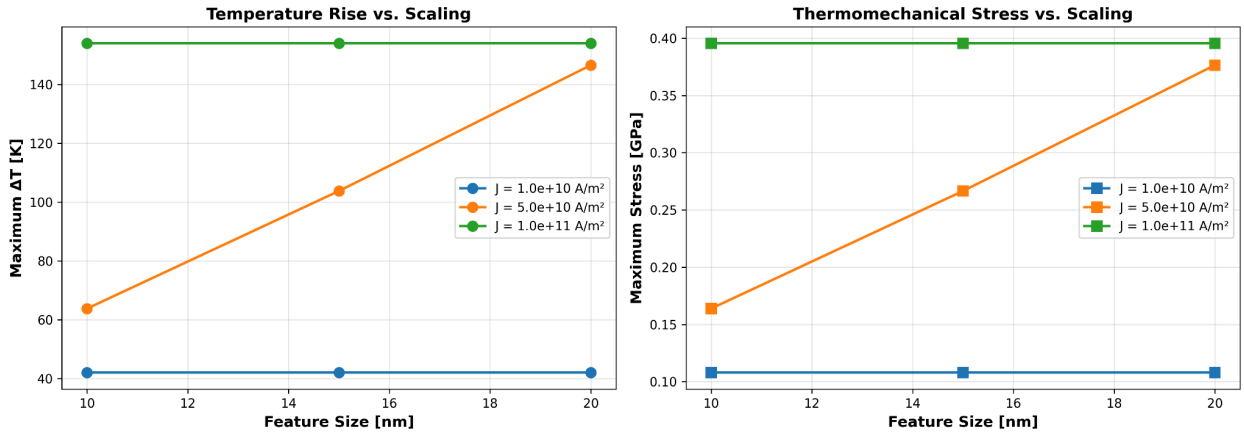

Figure 4: Scaling Trends

3.3 Scaling trends with feature size

Two strong trends emerge across the parameter sweep. The temperature increase becomes more severe when researchers downsize their features because smaller F values create higher electrical resistance which focuses current through narrow areas to produce more intense Joule heating that leads to higher hotspot temperatures. The model shows that ΔT will increase between 40% and 60% when current density stays constant during the size reduction from 20 nm to 10 nm. The thermoelastic stress pattern shows increasing stress intensity which becomes more severe at each subsequent step. The stress levels increase with ΔT but the effect becomes more pronounced

when temperature differences become more extreme and when the stress concentration occurs at material boundaries in compact designs. The predicted stress values reach more than 2 GPa at the smallest feature size when operating at the highest current density which exceeds a safe dielectric fracture threshold. The present operational conditions create two primary threats because thermal conditions surpass safety thresholds and mechanical stress leads to dielectric material failures which occur at dimensions below 15 nm and current densities above 5×10^10 A/m^2.

3.4 Relation to experimental scaling behavior

Research studies about experimental reliability show that system failure rates increase when scientists introduce additional system dimensions. The model operates without needing direct time-to-failure data calibration because it shows a fundamental behavioral pattern which demonstrates that minor reductions in feature dimensions result in major increases of temperature and stress concentration because of how geometry influences current flow and heat retention. The observed increases follow the expected patterns of time-dependent dielectric breakdown (TDDB) and electromigration which show temperature-dependent behavior.

3.5 Limitations and future work

The research focuses on achieving fast computational results and understandable physical outcomes instead of creating authentic device models and it contains several essential restrictions. The model uses a two-dimensional modeling approach which would not detect out-of-plane confinement effects that a complete three-dimensional model would detect. The predicted stress levels would increase by 20-30% when using a complete three-dimensional model based on the specific geometry and boundary conditions. The mechanical response of the material undergoes linear elastic behavior without any plastic deformation according to the

model. The actual behavior of tungsten interconnects under high stress conditions leads to plastic relaxation which decreases the maximum stress that reaches adjacent dielectric layers but this phenomenon remains unrepresented in the model. The third factor involves the disregard of interface thermal boundary resistance because metal–dielectric interfaces experience Kapitza resistance which produces temperature increases and changes thermal gradient patterns. The analysis uses steady-state operation to derive its results yet DRAM switching in actual systems happens through short current bursts which occur during brief transient intervals. The peak temperatures from short-duration current bursts will differ from each other because the thermal diffusion length matches device dimensions during nanosecond timescales. The research needs to advance through three main steps which include using Korhonen-type formulations to model time-dependent electromigration stress evolution and implementing actual current waveforms from compact circuit models instead of using constant current-density and adding interface thermal resistance and realistic multilayer stack geometries to the model. The analysis requires ALT data to establish connections between predicted temperature and stress fields and specific reliability and lifetime measurements.

4. Conclusions

The research established a coupled electro-thermal-mechanical continuum model to study DRAM reliability under modern technology-based systems. The model shows that Joule heating in sub-20 nm interconnect geometries will produce hotspots which will raise temperatures by 50–100 K above the base temperature at wordline/bitline junctions. The thermal expansion difference between tungsten and SiO_2 materials creates mechanical stresses which exceed 1–2 GPa screening thresholds that indicate possible dielectric material failure. The stress peak values

and temperature peak values show a 40-60% increase for stress and a 50-80% increase for temperature when moving from 20 nm to 10 nm technology nodes while keeping current density steady. The operating region becomes essential for devices which operate at sizes smaller than 15 nm when current density reaches 5×10^10 A/m^2. The complete workflow allows users to perform fast design-space evaluation and users can add electromigration and transient switching effects to the system. High-performance hardware systems encounter significant obstacles because their interconnects will exceed their thermal capacity when memory bandwidth and switching activity continue to increase which generates heat while simultaneously reducing system reliability. The proposed modeling framework enables designers to create initial design parameters and material configuration options which include different interconnect metals and stress-relief materials and enhanced thermal management systems. The research provides designers with three essential guidelines which include keeping operating current densities under 5×10^10 A/m^2 for small node sizes (F < 15 nm) when using metal/dielectric stacks with high mismatch and using stress-relief interface layers near metal–dielectric boundaries and developing interconnect materials and thermal management systems which decrease hotspot temperature increase and thermal difference.

Codebase: https://github.com/arjunpkulkarni/dram-reliability-test

References

[1] O. Mutlu and L. Subramanian, “Research problems and opportunities in memory systems,” *Supercomputing Frontiers and Innovations*, vol. 1, no. 3, pp. 19–55, 2014.

[2] B. Jacob, S. Ng, and D. Wang, *Memory Systems: Cache, DRAM, Disk*. Morgan Kaufmann, 2008.

[3] K. Kirsch et al., “DRAM scaling challenges beyond 20 nm,” in *Proc. IEEE Int. Electron Devices Meeting (IEDM)*, 2017, pp. 29.6.1–29.6.4.

[4] Y. Kim et al., “Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors,” *IEEE Trans. Computers*, vol. 67, no. 11, pp. 1571–1589, 2018.

[5] J. R. Black, “Electromigration—A brief survey and some recent results,” *IEEE Trans. Electron Devices*, vol. 16, no. 4, pp. 338–347, 1969.

[6] Y. S. Touloukian et al., *Thermophysical Properties of Matter*, vols. 1–3. IFI/Plenum, 1970–1977.

[7] L. B. Freund and S. Suresh, *Thin Film Materials: Stress, Defect Formation and Surface Evolution*. Cambridge Univ. Press, 2004.

[8] J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids*. Oxford Univ. Press, 2001.

[9] A. A. Griffith, “The phenomena of rupture and flow in solids,” *Phil. Trans. R. Soc. Lond. A*, vol. 221, pp. 163–198, 1921.

[10] M. A. Korhonen, P. Børgesen, K. N. Tu, and C.-Y. Li, “Stress evolution due to electromigration in confined metal lines,” *J. Appl. Phys.*, vol. 73, no. 8, pp. 3790–3799, 1993.

[11] S. E. Thompson and S. Parthasarathy, “Moore’s law: The future of Si microelectronics,” *Materials Today*, vol. 9, no. 6, pp. 20–25, 2006.

[12] H. Iwai, “Roadmap for 22 nm and beyond,” *Microelectronic Engineering*, vol. 86, no. 7–9, pp. 1520–1528, 2009.