

Chapter 2

3D Circuit Model Construction and Simulation

2.1 Introduction

[Chapter 1](#) shows the need to model the EM reliability of ULSI interconnects using 3D model at circuit layout level. In order to perform such modeling, a method to construct a complete 3D circuit model is necessary and this chapter will illustrate the construction and the corresponding transient electro-thermo-structural simulations for the EM reliability assessment of an IC.

Power and ground rails used to be the major concerns for the EM failure of the circuit due to the large, unidirectional currents in these lines. As the technology node goes down to 65 nm and below, the EM effects on signal wires and even in the logic cells themselves become significant [1]. To evaluate the EM reliability of a circuit, we should therefore look at both the power and the signal wires, as well as the global and the local interconnects.

To show a step by step approach, our illustration is carried out on a small circuit with two transistors, and only the transistor region (i.e., a simplified circuit structure that includes only the local interconnects with two metal levels) is considered. The circuit chosen is a simple inverter circuit and it is shown in [Fig. 2.1](#). The extension to the real circuit structure with the inclusion of other circuit components and the global interconnects up to the metal top will be discussed in detail in [Chap. 4](#).

2.2 Layout Extraction and 3D Model Construction

CMOS inverter is a basic building block in digital circuits [2]. Clock drivers made of inverter trees are prone to EM failure due to the large current demand on every cycle. The circuit layout of the inverter circuit shown in [Fig. 2.1](#) is drawn with Cadence using 0.18 μm technology based on the standard cell design of Global Foundry, as shown in [Fig. 2.2](#). The sizes of the PMOS and NMOS are kept small so as to reduce the size of the 3D model that is going to be built.

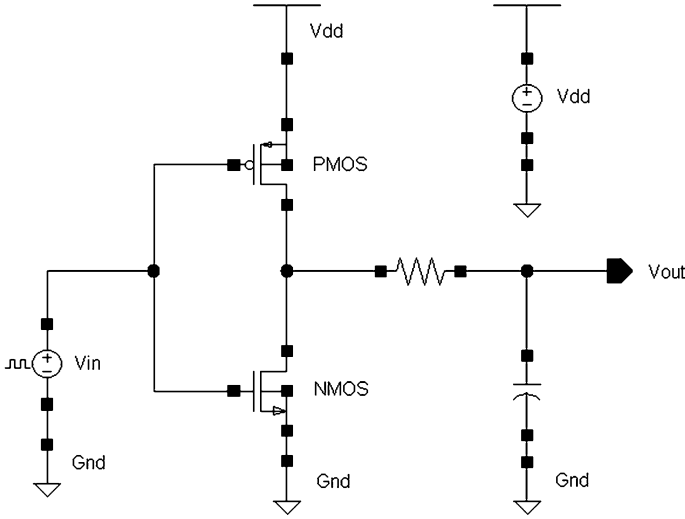


Fig. 2.1 Schematic of a simple inverter circuit

The 0.18 μm technology is chosen because the technology file is readily available and it will also be easier for the future experimental verification as the process technology is mature. As such, unnecessary unknown reliability issues can be avoided during the experimentation. The circuit designs using other technologies can also be used for the 3D model construction with the same method introduced here.

The Cadence GDSII file is first converted, layer by layer, into an ANSYS compatible file using third party software (e.g., LinkCAD). The thicknesses of each layer are specified during the conversion. Since our focus is on the EM reliability of interconnects, the transistors are assumed to be reliable. When the circuit is operating, the current flowing in the interconnects causes Joule heating and affects the temperature of the circuit, especially when the current spikes occur during switching. The focus of attention, the interconnect layers, together with the heat sources which are the transistor diffusion regions, are extracted and imported into ANSYS WORKBENCH. The converted layers are then combined and a 3D model is built [3]. As the layout tool provides only the basic interconnection structures, Ta barrier layer of 25 nm and SiN cap layer of 50 nm [4] are added around the interconnects, as shown in Fig. 2.3 [5, 6].

After importing the essential structures, additional layers that are present in the chip, such as the Si substrate, inter-level dielectric (ILD), passivation, as well as the packaging materials such as the plastic encapsulation, die attach and die pad (e.g., metal plate) as shown in Fig. 2.4 are added to the model, so as to represent a realistic chip condition. This will also enable us to perform chip-package interaction study, if needed. However, such interaction is outside the scope of this work.

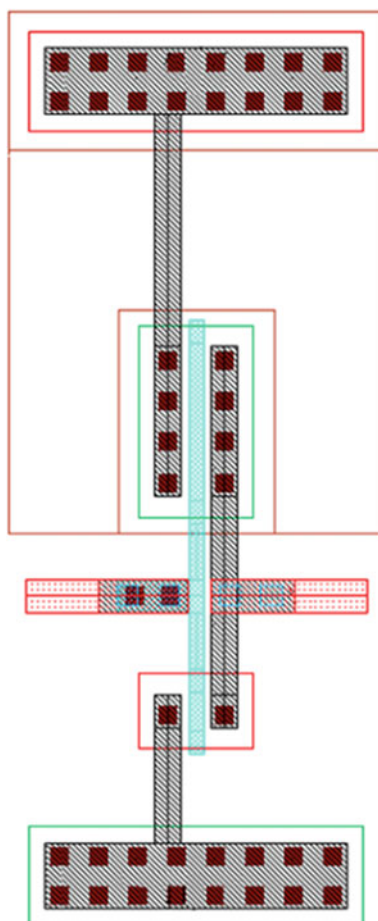


Fig. 2.2 Layout of a simple inverter circuit

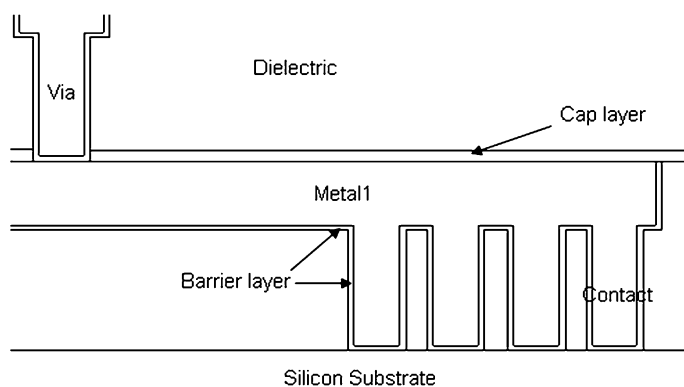


Fig. 2.3 Side view of the interconnects under study showing the barrier and cap layer (Reprinted from [6], with permission from Elsevier)

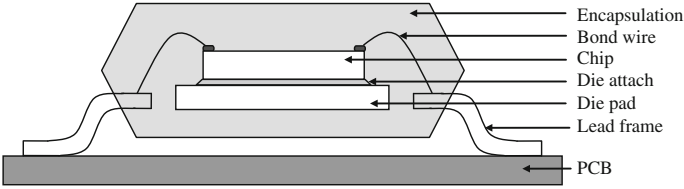


Fig. 2.4 Illustration of the plastic encapsulation used in a microelectronic device

To reduce the area effect on the circuit temperature, the substrate and the encapsulation are drawn to be much larger than the size of the inverter circuit. The complete 3D model with a size of $1,000 \times 1,000 \times 400 \mu\text{m}$ (i.e., the size of a die) is shown in Fig. 2.5. The inverter circuit is enclosed at the center just above the Si substrate, as indicated by the rectangular box in Fig. 2.5. A zoom in view of the inverter circuit is shown in Fig. 2.6. PS, PD, NS, and ND in Fig. 2.6 denote the PMOS source, PMOS drain, NMOS source, and NMOS drain respectively.

The geometrical parameter, the properties of the materials used, and the variation of thermal conductivity with temperature for the materials of interest are listed in Tables 2.1, 2.2, and 2.3 respectively [7–11]. The thermal conductivity of the encapsulation is assumed to be temperature independent in this case as the test temperature is below its glass transition temperature which is usually taken as 260°C [12].

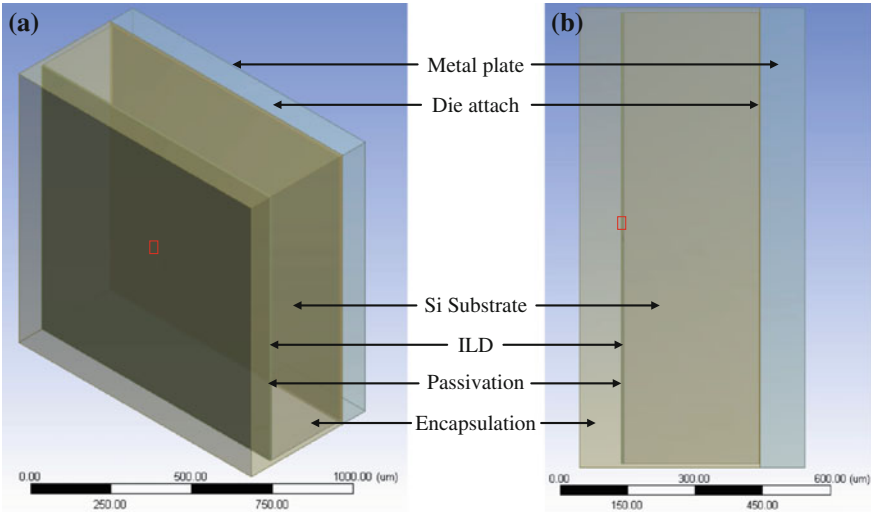


Fig. 2.5 **a** Isometric and **b** side view of the 3D model. The chip is modeled in a package with encapsulation and die attached to the lead frame represented by the metal plate shown in the figure. The model is in half transparent view for the ease of showing different layers

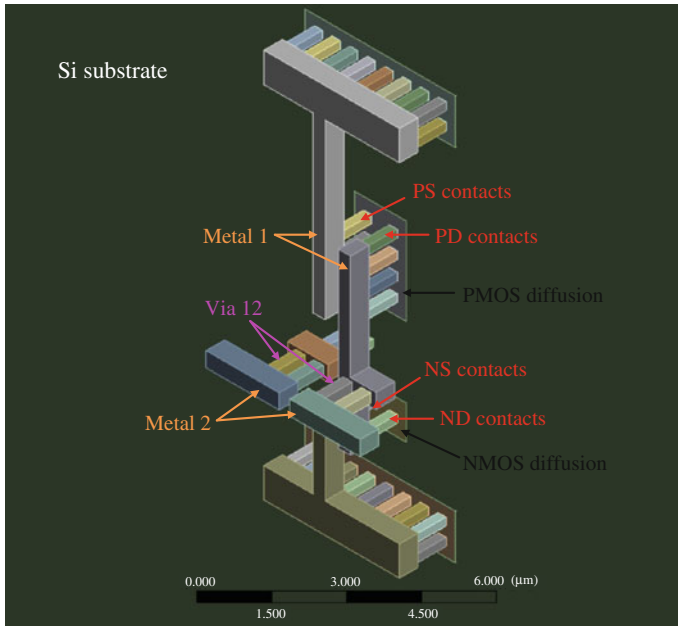


Fig. 2.6 Zoom in isometric view of the inverter circuit after removing the covered layers

Table 2.1 Geometry parameters of the model [8]

Feature	Size
Contact size	0.22 μm
Contact space	0.305 μm
Contact height	0.85 μm
Metal overlap contact	0.06 μm
Metal 1/Metal 2 width	0.34/0.42 μm
Metal 1/Metal 2 height	0.53 μm
Via 12 size	0.26 μm
Via 12 space	0.26 μm
Via 12 height	0.85 μm
PMOS diffusion region dimension (width $\times$ length)	2.46 $\times$ 1.46 μm
NMOS diffusion region dimension (width $\times$ length)	0.96 $\times$ 1.46 μm
Si substrate thickness	300 μm
Si substrate dimension (width $\times$ length)	1,000 $\times$ 1,000 μm
ILD thickness	3.10 μm
Passivation thickness	3.25 μm
Die attach thickness	2 μm
Metal plate thickness	100 μm
Encapsulation thickness	400 μm
Encapsulation dimension (width $\times$ length)	1,020 $\times$ 1,020 μm
Barrier layer thickness	25 nm
Cap layer thickness	50 nm

2.3 Transient Electro-Thermo-Structural Simulations and Atomic Flux Divergence Computation

After the construction of the 3D finite element circuit model, the next step is to perform the EM analysis. Different from the work reported in literature [13–16] which used DC analysis, the simulation of our 3D circuit model is performed under transient condition so as to account for the possible temperature and stress effects due to the change in magnitude and/or direction of the current flow. Coupled thermal-electric and structural-thermal simulations are performed to calculate the AFDs in the interconnects due to different driving forces, and these values are used to evaluate the EM reliability of the interconnects.

2.3.1 Transient Thermal-Electric Analysis

The transient thermal-electric analysis is performed to study the effect of the electrical loads on the temperature and current density distributions. For an accurate simulation, the boundary conditions and the electrical loads should be properly set.

2.3.1.1 Boundary Conditions

The running speed, the cooling method, and the operation modes of the transistors determine the temperature of an IC. The desktop processors usually run at a temperature between 70 and 90 °C [17]. In this work, a worst-case temperature of 90 °C is used to represent the interconnect heating in a chip under operation [18]. Other temperatures can be used depending on the application of the circuits. The metal base plate is treated as the heat sink since its thermal conductivity is more than 100 times better than that of the encapsulation material as can be seen in Table 2.2, and thus it is the major heat dissipation path for the model. A convection heat transfer coefficient h of $20 \text{ W/m}^2 \cdot ^\circ\text{C}$ [19] is used to simulate the 90 °C ambient condition. A constant h is used for simplicity as the convection has minor effect to the model temperature in this work as verified by our simulation where the model temperature is found to remain unchanged when the value of h varies from 10 to $40 \text{ W/m}^2 \cdot ^\circ\text{C}$. The boundary conditions of the thermal-electric simulation used in the model are shown in Fig. 2.7.

2.3.1.2 Electrical Loads

The instantaneous source and drain currents flowing in the interconnects, together with the voltages across the interconnects, are the causes of the heat up of the

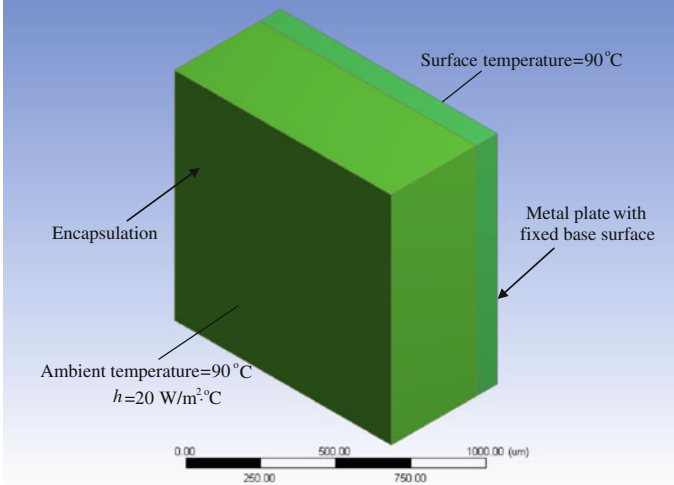
Table 2.2 Material properties used in the model [7, 9, 10]

Materials	Properties						
	Young's Modulus (MPa)	Poisson's Ratio	Density (kg/m ³)	Thermal Expansion (1/°C)	Thermal Conductivity (W/m·°C)	Specific Heat (J/kg·°C)	Resistivity (Ohm·m)
Cu	1.10×10^5	0.34	8,300	1.80×10^{-5}	403	385	1.72×10^{-8}
Si	1.30×10^5	0.28	2,330	2.60×10^{-6}	149	700	1.00×10^{17}
Substrate							
As-doped Si							9.90×10^{-3}
B-doped Si							2.81×10^{-2}
Heavily doped Si							1.00×10^{-5}
Si							
SiO ₂	7.14×10^4	0.16	2,200	6.80×10^{-7}	1.38	1000	1.00×10^{17}
SiN	2.20×10^5	0.27	3,100	3.20×10^{-6}	30	700	1.00×10^{13}
Ta	1.86×10^5	0.34	16,690	6.30×10^{-6}	57.50	140	1.31×10^{-7}
Polyimide	3,100	0.33	1,430	5.00×10^{-5}	0.18	1100	1.00×10^{15}
96 % Alumina	276	0.25	3,965	7.10×10^{-6}	20.90	779	1.00×10^{13}

Table 2.3 Temperature dependency of thermal conductivity for some materials [11]

Materials	Thermal conductivity (W/m · °C)					
	0 °C	27 °C	127 °C	227 °C	327 °C	427 °C
Cu	403	395	390	380	370	365
Si Substrate	149	148	99	76	62	51
SiO ₂	1.38	1.38	1.51	1.61	1.75	1.92

Reprinted from He and Tan [5], with permission from Elsevier

**Fig. 2.7** Boundary conditions of the thermal-electric simulation

interconnects. These currents and voltages are treated as the electrical loads in the ANSYS thermal-electric simulation, and their values are determined from the circuit design tool, Cadence.

The power consumed in the inverter circuit includes the dynamic power due to the charging and discharging of the capacitors, and the short circuit power due to the direct current flow from the power source to the ground during switching when both transistors are on [20]. The diode leakage power is assumed to be negligible and is ignored in this simulation. To simulate the power consumption in the inverter, both the charging and discharging currents and the short circuit current should be considered.

When PMOS is on, the path of the current flow is from the V_{dd} line to the source of PMOS, and then from the drain of PMOS to the output line. For the NMOS-on condition, the current flows from the output line through the NMOS to the ground. The current flow paths and the name of the four current sources are listed below:

1. the current flow from the V_{dd} line to the source of PMOS: I_{ps} ,
2. the current flow from the drain of PMOS to the output line: I_{pd} ,
3. the current flow from the output line to the drain of NMOS: I_{nd} ,

4. the current flow from the source of NMOS to the ground: I_{ns} .

The above charging and discharging currents are represented by the gradual increase or decrease of the current waveforms. The short circuit currents are represented by the current spikes during switching and they quickly drop to zero after switching. The currents in the gates of the two transistors are assumed to be negligible.

The root-mean-square (RMS) current density is usually used in the EM simulations performed under DC or low frequency pulsed DC condition. The study in [21] showed that, the steady-state simulation using the RMS current density of 1.5 MA/cm^2 gives a small temperature difference of around 4 % as compared with the transient simulation using an unipolar pulsed DC load at 1 MHz. With a pulsed DC load, EM occurs during the switch-on (i.e., DC) period and nothing or stress relaxation may occur during the switch-off (i.e., no current flow) period [22]. At low frequency of no more than 1 MHz, the metal line can respond thermally to the pulses and thus has a higher temperature during the switch-on period than during the switch-off period. Relaxation of the EM induced stress does not occur during the switch-off period as the temperature of the metal line is low. Therefore the RMS values can be used in the low frequency regime. However, at higher frequencies, the switching speed is faster than the thermal response time of a typical thin-film conductor (i.e., a few μs), stress relaxation occurs as the metallization cannot respond thermally to the pulses [23]. The stress relaxation during the switch-off period results in an enhanced EM lifetime and this effect is overlooked if the RMS value is used. Under an AC operation condition, besides the stress relaxation effect at high frequency, the bi-directional current (e.g., at the output line of the inverter) further affect the stress state in the metal line. It is found in [24] that the EM damage by the positive (forward) current stress can be partially healed by the following negative (reverse) current stress. Therefore it is not suitable to use the RMS values under the AC condition with a frequency higher than 1 MHz, and the instantaneous electrical loads under the AC operation condition are used in our simulation.

The voltage and current pulses at selected circuit nodes (i.e., the source and drain of the two transistors, the V_{dd} and ground lines, the input and output lines) are extracted from Cadence, including both the charging and discharging currents and the current spikes during switching. These current and voltage values are used as the inputs to the 3D circuit model in ANSYS.

In the delay analysis of the CMOS inverter driving a RC load in [25], the load resistance varies from 10 to 1 k Ω , and the load capacitance is in the range of 0.01–1 pF. Here an extreme case of 1 k Ω and 1 pF under an operation frequency of 100 MHz is used as an example. Also, a large load and low frequency are used so that the power consumption of the circuit is larger and the temperature change within one operation cycle is more obvious. Furthermore, the difference between the temperature responses to the rapid current change (e.g. current spike) and the gradual current change (e.g., slow charging or discharging) can be clearer. This non-typical condition is for the validation of the 3D model construction and simulation. Other typical loading and operating frequencies can be used in

accordance to the application in a real circuit under the realistic circuit operation condition.

When high frequency AC current flows in a conductor, it tends to redistribute itself within a conductor with the current density being largest near the surface of the conductor, and decreasing at greater depths. The electric current concentrates at the “skin” of the conductor and this phenomenon is called the “skin effect”. The depth at which the current density has fallen to $1/e$ (about 0.37) of its original value at the conductor surface is called the “skin depth” and is calculated as [26],

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} \quad (2.1)$$

where δ is the skin depth, ω is the angular frequency of current ($2\pi \times$ frequency), and μ is the absolute magnetic permeability of the conductor.

When the skin effect occurs, the effective cross-section of the conductor decrease and its effective resistance increases, especially at higher frequencies where the skin depth is smaller. The increase in the effective resistance of the conductor increases the Joule heating and hence enhances EM. The increased current density at the surface further increases the surface temperature and aggravates the EM process.

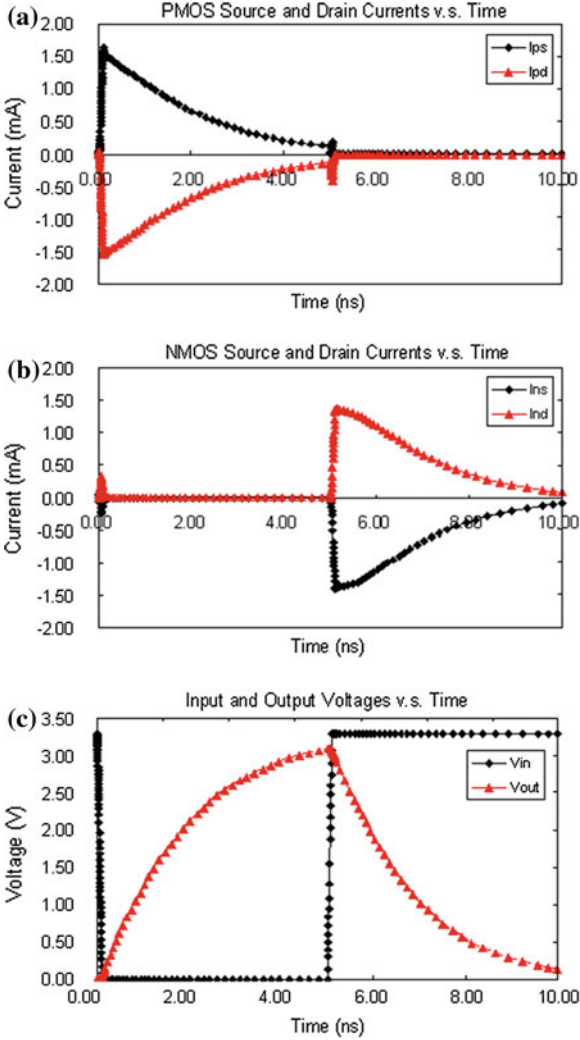
At the operating frequency of 100 MHz is used in this case, the skin depth calculated based on Eq. (2.1) is $6.60 \mu\text{m}$, and it is larger than the thicknesses of Metal 1 ($0.53 \mu\text{m}$), Metal 2 ($0.53 \mu\text{m}$), Via 12 ($0.85 \mu\text{m}$), and the contacts ($0.75 \mu\text{m}$), and the skin effect is ignored in this model.

The electrical loads of the inverter circuit within one load cycle (i.e., one period of 10 ns) under the condition specified above are shown in Fig. 2.8. We can see that the PMOS source and drain current waveforms show the same pattern but in opposite direction as expected. The same trend is observed for the NMOS current waveforms. The output voltage does not switch instantaneously with the input voltage due to the presence of the RC load. The threshold voltage V_t of the PMOS and NMOS are -0.495 and 0.482 V respectively, which are obtained from our calculation using Cadence.

As it is not possible to include all the thousands of data points in Fig. 2.8 into the 3D circuit simulation, the data set should be simplified. As can be seen in Fig. 2.8, the electrical loads within one load cycle can be divided into four stages: the first switching stage, the first operation/off stage, the second switching stage, and the second operation/off stage. These four stages are called Stage 1–4 respectively.

For I_{ps} and I_{pd} , there is an only-rise/fall current spike at Stage 1, a gradual current change waveform (i.e., transistor in operation, slow charging of the capacitor) at Stage 2, a rise-and-fall current spike at Stage 3, and a constant current waveform (i.e., transistor is off, almost zero current) at Stage 4. Similarly, for I_{ns} and I_{nd} , there is a rise-and-fall current spike at Stage 1, a constant current waveform (i.e., transistor is off, almost zero current) at Stage 2, an only-rise/fall current spike at

Fig. 2.8 Electrical loads of the inverter circuit within one load cycle (10 ns) showing **a** the PMOS source and drain currents **b** the NMOS source and drain currents and **c** the input and output voltages



Stage 3, and a gradual current change waveform (i.e., transistor in operation, slow discharging of the capacitor) at Stage 4. For V_{in} , there is an only-rise/fall voltage spike at Stage 1 and 3, and a constant voltage waveform at Stage 2 and 4. For V_{out} , there is a gradual voltage increase waveform at Stage 1 and 2, and a gradual voltage decrease waveform at Stage 3 and 4.

The selection of data points depends on the shape of the current and voltage waveforms used in the simulation. For example, we should select more data points at the region with more significant and/or irregular changes. More data points should be chosen if particular accuracy is required at a specific time interval.

In this case, the gradual current or voltage change waveforms need around 5 data points to represent their shapes. The number of data points selected can

be changed depending on the accuracy requirement and the PC computation capability. Here 5 points are used so that there can be a gradual change in current or voltage at around 1 ns interval. The only-rise/fall current or voltage spikes need 5 data points: 1 at the beginning, 1 at the peak, 3 on the rise/fall slope. The rise-and-fall current spikes need 8 data points: 1 at the beginning, 1 at the peak, 3 on the rise-slope and 3 on the fall-slope. The constant current or voltage waveforms need only 2 data points: 1 at the beginning and 1 at the end. As the current and voltage waveforms occur simultaneously, there are 6 waveforms with different shapes in each stage. When the waveforms within the same stage have different shapes, the number of data points selected is determined by the waveform that needs more points to represent its shape.

Therefore, at Stage 1, the only-rise/fall current spikes for I_{ps} and I_{pd} need 5 data points; the rise-and-fall current spikes for I_{ns} and I_{nd} need 8 data points; the voltage spike for V_{in} need 5 data points; and the gradual voltage increase waveform for V_{out} need 5 data points. Therefore 8 data points are selected at Stage 1. At Stage 2, the gradual current change waveforms for I_{ps} and I_{pd} need 5 data points; the constant current waveforms for I_{ns} and I_{nd} and the constant voltage waveform for V_{in} need 2 data points; and the gradual voltage increase waveform for V_{out} need 5 data points. Therefore 5 data points are selected at Stage 2. At Stage 3, the rise-and-fall current spikes for I_{ps} and I_{pd} need 8 data points; the only-rise/fall current spikes for I_{ns} and I_{nd} need 5 data points; the voltage spike for V_{in} need 5 data points; and the gradual voltage decrease waveform for V_{out} need 5 data points. Therefore 8 data points are selected at Stage 3. At Stage 4, the constant current waveforms for I_{ps} and I_{pd} need 2 data points; the gradual current change waveforms for I_{ns} and I_{nd} need 5 data points; the constant voltage waveform for V_{in} need 2 data points; and the gradual voltage decrease waveform for V_{out} need 5 data points. Therefore 5 data points are selected at Stage 4.

The data point selection at Stages 1 and 4 for I_{nd} and Stage 2 and 3 for I_{ps} is shown in Fig. 2.9. The time scale at Stage 1 and 3 (i.e., the current spike stages) in Fig. 2.9 are exaggerated for better illustration purpose. The data point selection for all the current and voltage waveforms at the four stages is summarized in Table 2.4.

In total, 26 data points are selected to present the shapes of the current and voltage waveforms in Fig. 2.8. The simplified current and voltage waveforms within one load cycle are shown in Fig. 2.10. Each data point corresponds to one load step in the ANSYS simulation. A do-loop is used to repeat the 26 load steps for multiple cycles to represent a continuous circuit operation.

2.3.1.3 Application of the Electrical Loads

The locations for applying the current and voltage loads in the 3D circuit model correspond to the data extraction nodes in Cadence, and they are shown in Fig. 2.11. The direction of the current flow is represented by the + and – sign in the simulation.

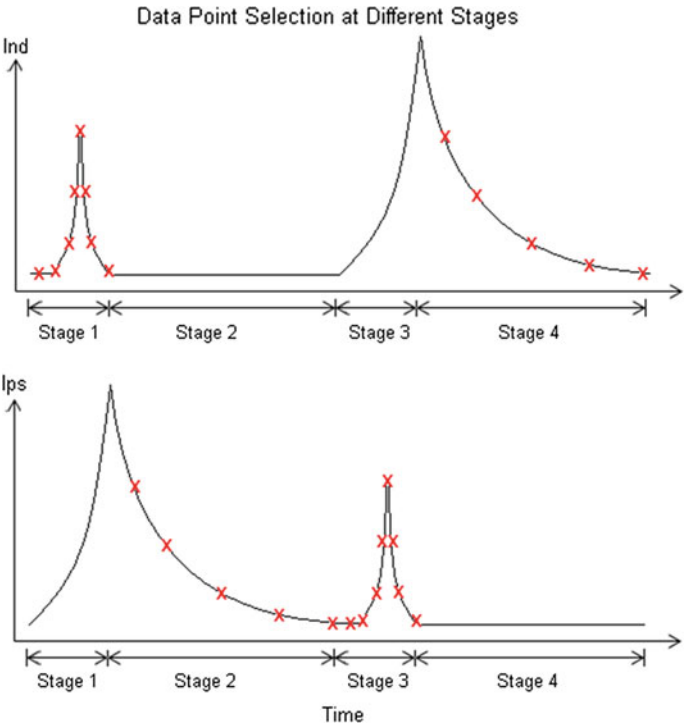


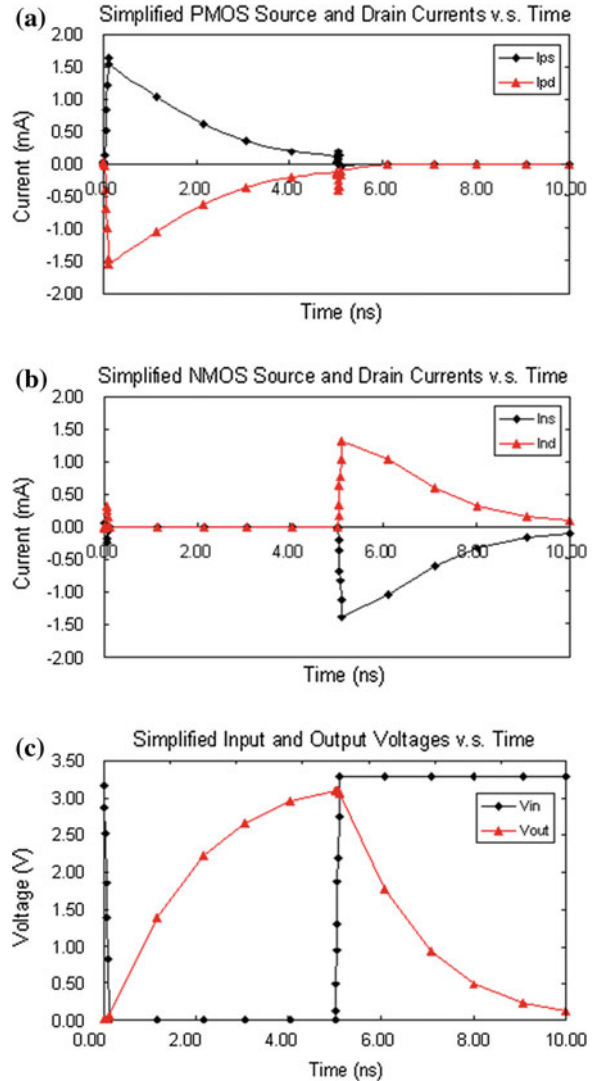
Fig. 2.9 Data point selection at different stages (I_{nd} and I_{ps}), not to scale

Table 2.4 Current and voltage waveforms and the data point selection at different stages. The number in the bracket is the data points selected at this stage for this waveform

Stage	Current waveforms		Voltage waveforms		Data points
	I_{ps} and I_{pd}	I_{ns} and I_{nd}	V_{in}	V_{out}	
1	Only-rise/fall spike (5)	Rise-and-fall spike (8)	Only-fall spike (5)	Gradual increase (5)	8
2	Gradual change (5)	Constant (2)	Constant (2)	Gradual increase (5)	5
3	Rise-and-fall spike (8)	Only-rise/fall spike (5)	Only-rise spike (5)	Gradual decrease (5)	8
4	Constant (2)	Gradual change (5)	Constant (2)	Gradual decrease (5)	5

After the application of the boundary conditions and the electrical loads, the transient thermal-electric simulation is performed and the results are discussed in [Sects. 2.4.1](#) and [2.4.2](#).

Fig. 2.10 Simplified electrical loads of the inverter circuit within one load cycle (10 ns) showing **a** the PMOS source and drain currents **b** the NMOS source and drain currents and **c** the input and output voltages



2.3.2 Transient Structural-Thermal Analysis

Besides the intrinsic stress that is already present at the beginning of the simulation, when the current and voltage loads are applied to the model, there is a difference in the interconnect temperature at different locations, and the thermal expansivity mismatch among the materials in an interconnect structure causes thermally induced thermo-mechanical stress. The transient structural-thermal analysis is needed to study the thermo-mechanical stress distribution in the interconnects.

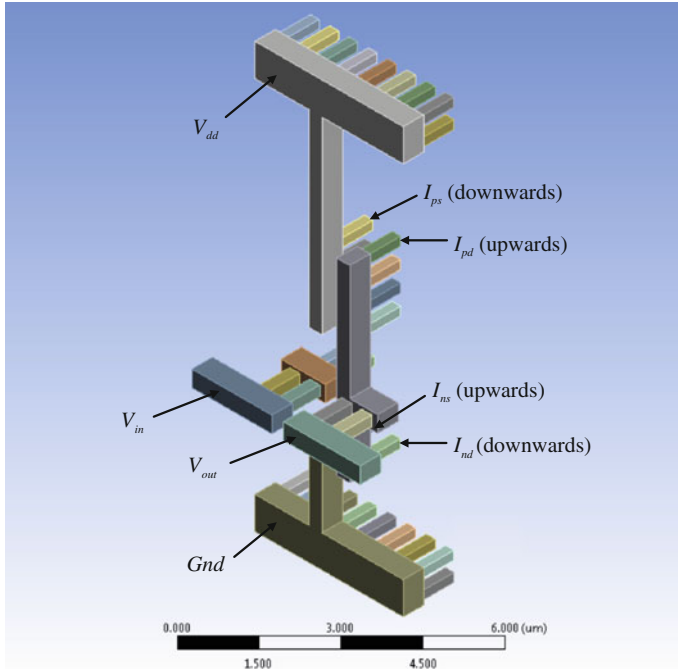


Fig. 2.11 Application of the current and voltage loads to the 3D circuit model in ANSYS, showing the current flow direction during circuit operation. Only the interconnects are shown for clarity

2.3.2.1 Boundary Conditions

All the nodes on the base surface of the model are assumed to be fixed as they are mounted on the circuit board. The stress free temperature (SFT) of the Cu dual-damascene structure is usually between 350 and 400 °C [27, 28], depending on the final annealing temperature of the sample. In this work, a SFT of 350 °C is used.

2.3.2.2 Thermal Loads

The temperatures of all the nodes of the model at each load step are retrieved from the results of the transient thermal-electric simulation. These temperature values, together with the time of their occurrence, are used as the thermal loads in the transient structural-thermal simulation.

After the application of the boundary conditions and the thermal load, the transient structural-thermal simulation is performed and the results are discussed in Sect. 2.4.4.

2.3.3 Application of Submodeling

A fine mesh is needed to achieve an accurate simulation result. The original full model is too large and thus very time and resource consuming for the transient analysis if a fine mesh is applied to the entire model. As a result, submodeling is used in this simulation.

Submodeling, also known as cut-boundary displacement method or specified boundary displacement method, cuts through the global model. A fine mesh is done at the “cut out” region, while the global model can be meshed much coarser. Mesh refinement is achieved at the cut out sub-model region and thus we are able to obtain an accurate simulation result while eliminate the redundant effort on solving the region that is not the focus of attention. The main disadvantage of the submodeling method is that the simulation needs to be performed twice, first for the global model and then for the sub-model, and thus complicated the process.

When submodeling is applied, a cut boundary (i.e., the boundary that cuts through the coarse model) is specified, the displacement and/or temperature results extracted from the cut boundary of the coarse model are applied as the boundary conditions of the sub-model. According to St. Venant’s principle [29], if an actual distribution of forces is replaced by a statically equivalent system, the distribution of the stress and strain is changed only near the regions of load application. This implies that, acceptable results can be obtained in the sub-model if the cut boundaries of the sub-model are adequately far away from the stress concentration, in view of the fact that the stress concentration effects are localized around the concentration sites [30].

A simple steady-state simulation is carried out on the full model to determine where the cut boundary should be, i.e., the size of the sub-model. With the same setup for the boundary conditions in Sect. 2.3.1.1, the first current spike is used as the input for the simulation. This current spike is used because it is one of the worst-case values which can result in a higher temperature than that using the normal operation current in the circuit. The resultant worst-case temperature is sufficient for determining the cut boundary location.

From Fig. 2.12, it can be seen that the temperature is almost constant at around 5 μm away from the model center in the x direction and 10 μm in the y direction, as indicated by the rectangular box. Hence it is acceptable to place the cut boundary there. A $15 \times 20 \times 12 \mu\text{m}$ sub-model centered with the inverter circuit is cut out from the original $1,000 \times 1,000 \times 400 \mu\text{m}$ full model as shown in Fig. 2.13.

The complexity of the model limits the application of the hexahedral mesh. Wang et al. [31] showed that the quadratic tetrahedral shaped element can produce as accurate result as the linear hexahedral element, and hence a quadratic tetrahedral mesh using 20-node coupled-field solid element SOLID98 is used in this work. The quadratic tetrahedral mesh is applied to both the full model and the sub-model. A fine mesh with a size of 0.50 μm is used for the interconnects in the full model, and an even finer mesh is used in the sub-model. A transient analysis using the electrical loads in Fig. 2.10 is conducted to determine the mesh size to be used

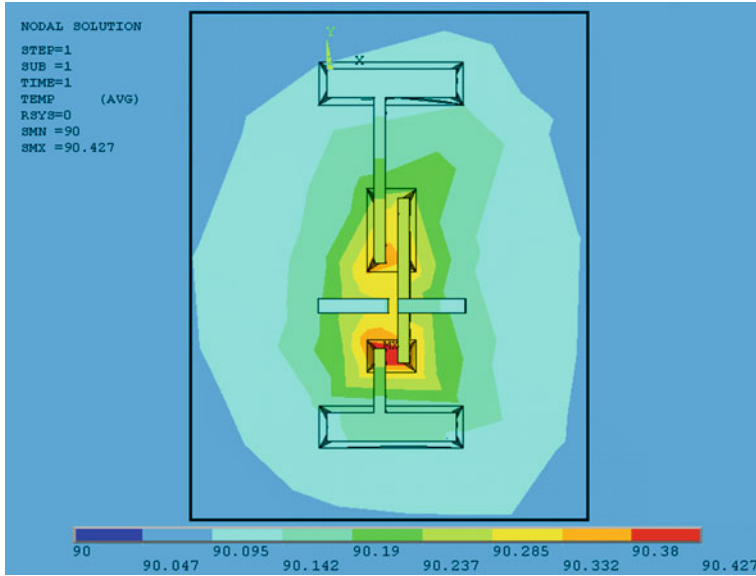


Fig. 2.12 Plane view of the steady-state temperature distribution of the full model using the worst-case current. The *box* indicates the margin where the temperature becomes almost constant

in the sub-model and the simulation results at the end of one load cycle (i.e., 10 ns) are shown in Fig. 2.14.

From Fig. 2.14, it is clear that a fine mesh of $0.20\ \mu\text{m}$ is adequate for the sub-model as the temperature and the stress curves both converge (i.e., the curve becomes flat) at $0.20\ \mu\text{m}$. Similarly, the barrier layer can be meshed at $0.10\ \mu\text{m}$ without the loss of accuracy. The coarse mesh for the full model and fine mesh for the sub-model are shown in Fig. 2.15.

The outer surfaces of the sub-model are treated as the cut boundary. According to the transient submodeling analysis technique introduced by Wang et al. [32], the simulation results of the full model at each load step are recorded, and the solution of the first load step is retrieved and applied to the cut boundary of the sub-model as the boundary conditions. The sub-model is then solved with the current and the voltage loads and the applied boundary conditions from the first load step. This is repeated for the remaining load steps.

2.3.4 Computation of Atomic Flux Divergences

Finite element analysis using the AFD approach is a common practice in 3D EM modeling. This approach is used as there is a direct relationship between the EM

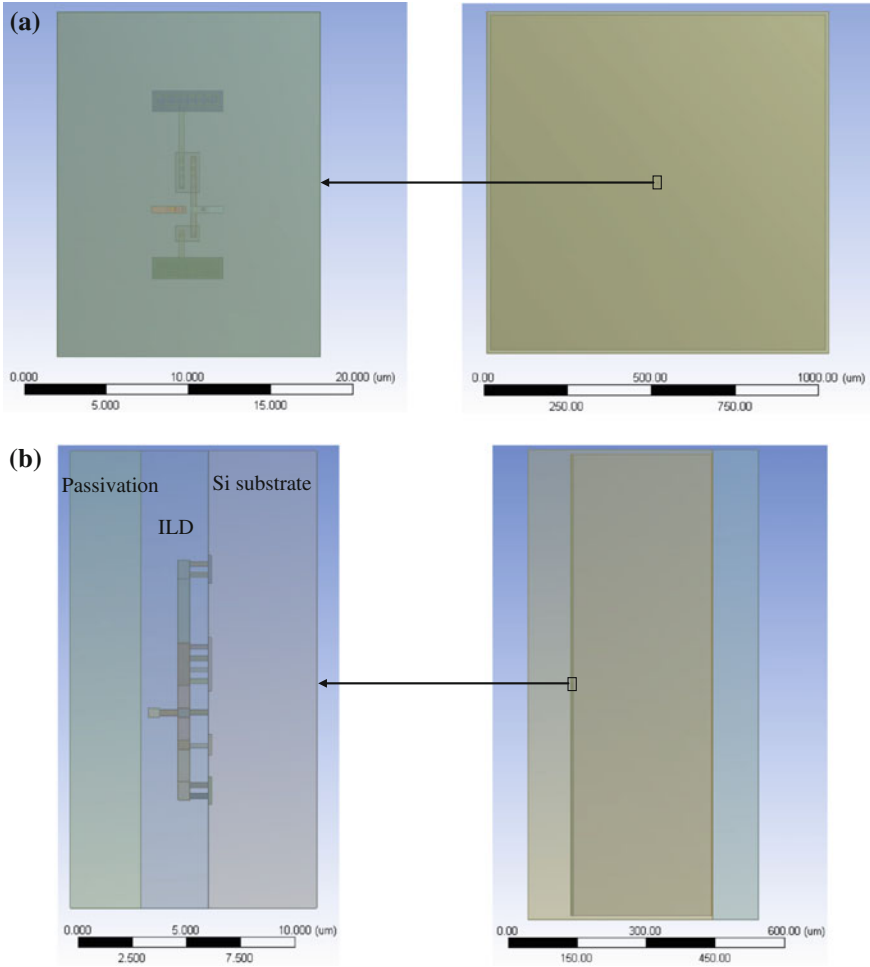


Fig. 2.13 The cut out of the sub-model (*left*) from the full model (*right*), models shown in **a** plane and **b** side views

failure time and the AFD in a small interconnect segment of length Δl for a time period of Δt [33].

$$t_f = \frac{V_c}{\Omega' \Delta l p_w p_t \nabla \cdot J} \quad (2.2)$$

where t_f is the time-to-failure, V_c is the critical volume of mass depletion or accumulation required for failure to occur, Ω' is the atomic volume, p_w and p_t are the effective diffusion path width and thickness respectively, and $\nabla \cdot J$ is the total atomic flux divergence.

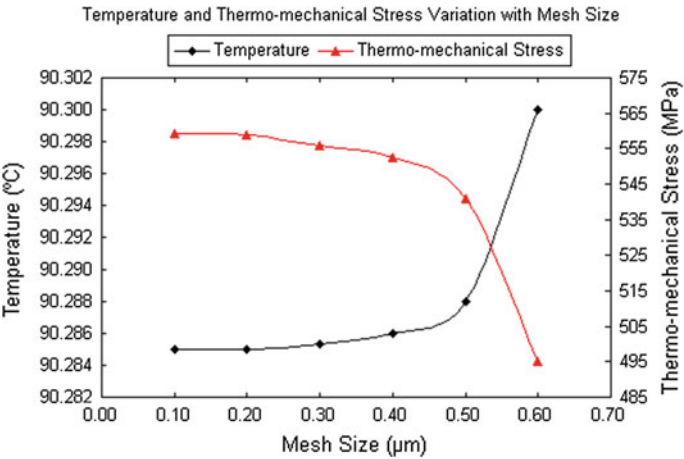


Fig. 2.14 Temperature and thermo-mechanical stress variation with mesh size showing convergence achieves at 0.20 μm

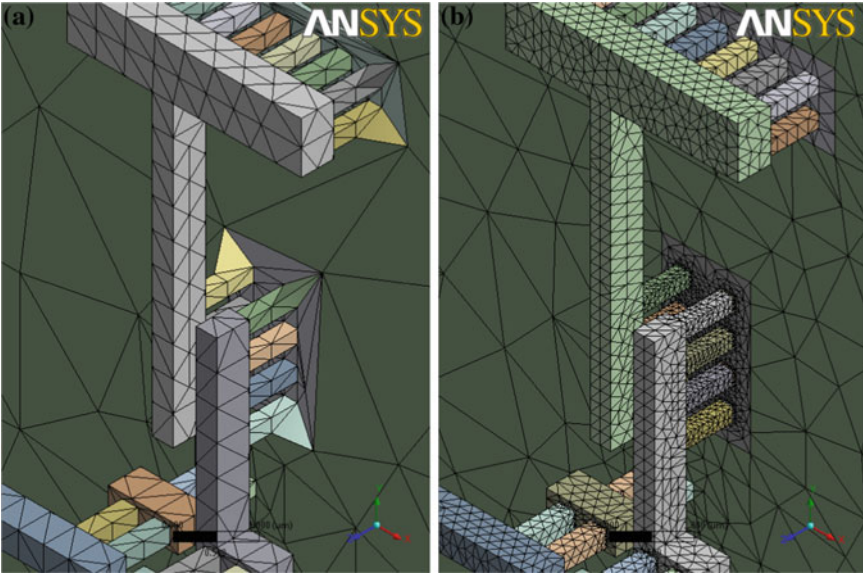


Fig. 2.15 Tetrahedral mesh at the interconnect region showing **a** the coarse mesh for the full model and **b** the fine mesh for the sub-model

From Eq. (2.2), it is obvious that the EM failure time of the interconnects is inversely proportional to the AFD. Void is assumed to nucleate at the location with the highest positive AFD and grows irreversibly. Experimental observation in [34] showed that the void was indeed found at the location of the maximum AFD with excellent agreement, indicating the accuracy of this approach.

The equations used in our model are based on the work by Dalleau et al. [13], and we extend their model from line-via level to circuit layout level. There are three dominant driving forces for the AFD, namely the electron wind force, the temperature gradient induced driving force, and the thermo-mechanical stress gradient induced driving force. The atomic fluxes due to the three driving forces were modeled using Eqs. (2.3)–(2.5) respectively [13],

$$\vec{J}_{\text{EWF}} = \frac{N}{k_B T} e Z^* j \rho D_0 \exp\left(-\frac{E_a}{k_B T}\right) \quad (2.3)$$

$$\vec{J}_T = -\frac{N Q^* D_0}{k_B T^2} \exp\left(-\frac{E_a}{k_B T}\right) \nabla T \quad (2.4)$$

$$\vec{J}_S = \frac{N \Omega' D_0}{k_B T} \exp\left(-\frac{E_a}{k_B T}\right) \nabla \sigma_H \quad (2.5)$$

where J_{EWF} , J_T and J_S are the atomic fluxes due to electron wind force, temperature gradient induced driving force and thermo-mechanical stress gradient induced driving force respectively, N is the atomic concentration, k_B is the Boltzmann constant, T is the temperature of the metal line, $e Z^*$ is the effective charge of the diffusing species, j is the current density, ρ is the electrical resistivity which is dependent on temperature given as $\rho = \rho_0 (1 + \alpha(T - T_0))$, α is the temperature coefficient of resistivity, ρ_0 is the electrical resistivity at temperature T_0 , D_0 is the prefactor of the self-diffusion coefficient of the metal line, E_a is the activation energy for self-diffusion of the metal line, Q^* is the heat of transport of the metal line, σ_H is the local hydrostatic stress calculated using the average of the hydrostatic stress values in the interconnect in the x, y, z direction as $\sigma_H = (\sigma_{xx} + \sigma_{yy} + \sigma_{zz})/3$.

After obtaining the atomic fluxes, the divergences due to electron wind force (AFD_EWF), temperature gradient induced driving force (AFD_T), and thermo-mechanical stress gradient induced driving force (AFD_S) were calculated using Eqs. (2.6)–(2.8) respectively [13]:

$$\nabla \cdot \vec{J}_{\text{EWF}} = \left(\frac{E_a}{k_B T^2} - \frac{1}{T} + \alpha \frac{\rho_0}{\rho} \right) \cdot \vec{J}_{\text{EWF}} \nabla T \quad (2.6)$$

$$\nabla \cdot \vec{J}_T = \left(\frac{E_a}{k_B T^2} - \frac{3}{T} + \alpha \frac{\rho_0}{\rho} \right) \cdot \vec{J}_T \nabla T + \frac{N Q^* D_0}{3 k_B^3 T^3} \vec{j}^2 \rho^2 e^2 \exp\left(-\frac{E_a}{k_B T}\right) \quad (2.7)$$

$$\begin{aligned} \nabla \cdot \vec{J}_S = & \left(\frac{E_a}{k_B T^2} - \frac{1}{T} \right) \cdot \vec{J}_S \nabla T + \frac{2EN\Omega' D_0 \alpha'}{3(1-\nu)k_B T} \exp\left(-\frac{E_a}{k_B T}\right) \left(\frac{1}{T} - \alpha \frac{\rho_0}{\rho} \right) \nabla^2 T \\ & + \frac{2EN\Omega' D_0 \alpha'}{3(1-\nu)k_B T} \exp\left(-\frac{E_a}{k_B T}\right) \frac{\vec{j}^2 \rho^2 e^2}{3k_B^2 T} \end{aligned} \quad (2.8)$$

where E , α' and ν are the Young's Modulus, the thermal expansion coefficient, and the Poisson ratio of the metal line respectively.

To compute the AFDs in the interconnects, the thermal-electric simulation is first carried out using the boundary conditions and the electrical loads as shown in Sects. 2.3.1.1 and 2.3.1.2. The nodal temperatures extracted from the thermal-electric simulation are then used as the thermal loads for the structural-thermal simulation. The structural-thermal simulation is performed based on the applied thermal loads and the boundary conditions in Sect. 2.3.2.1. Submodeling technique is used in both simulations. The thermo-mechanical stress gradient of the model is calculated based on the thermo-mechanical stresses extracted from the structural-thermal simulation. The thermo-mechanical stress gradient, together with the current density, temperature, and temperature gradient extracted from the thermal-electric simulation, are used to compute the atomic fluxes and the flux divergences due to electron wind force, temperature gradient induced driving force, and thermo-mechanical stress gradient induced driving force, using Eq. (2.3)–(2.8). The flowchart for the computation of the AFDs in ANSYS is shown in Fig. 2.16. The total AFD in the interconnects is the sum of the three AFDs from their respective driving forces and the results obtained are discussed in Sect. 2.4.5.

The simulations are carried out using an Intel Quad core computer (Q9400@ 2.66 GHz) with 4G RAM. The computation time and result file size of each transient simulation for one load cycle (10 ns, 26 load steps) are listed in Table 2.5 [5].

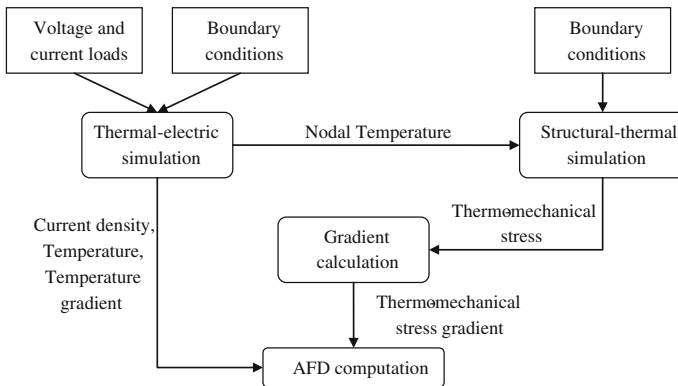


Fig. 2.16 Flowchart for the computation of the atomic flux divergences

Table 2.5 The computation time and result file size of each transient simulation for one load cycle

Simulation	Computation time (min)	Result file size (GB)
Thermal-electric simulation	70	9.81
Structural-thermal simulation	156	9.52
Thermo-mechanical stress calculation & AFD computation	144	6.60
Total	370	25.93

It can be seen from the listed values that the overall simulation takes about 6 h running time and the result file is nearly 26 GB.

Due to the complexity of the problem and the constraint of time and resource, the current simulation is focusing on the void nucleation location instead of the dynamics of void growth. This is justifiable due to the following two reasons. Firstly, recent simulation and experimental results by Li et al. [35] showed that the void nucleation time occupies the largest portion of the EM lifetime, and thus reducing the void nucleation time can effectively enhance the EM lifetime. Secondly, it is always important to identify the weakest site or the damage nucleation site so that one can strengthen these sites for the “design-in” reliability of a device or system.

2.4 Simulation Results and Discussions

2.4.1 Current Density and Temperature Distributions

The current density and temperature distributions of the 3D model at any load step can be viewed after the thermal-electric simulation.

Figures 2.17 and 2.18 show the current density and temperature distributions of the interconnects of the inverter model under study after one and ten load cycles (i.e. 10 and 100 ns) respectively. The figures on the right of Fig. 2.17a are the zoom-in views at the PMOS and NMOS contact regions. For a better illustration, the zoom-in views are drawn using their own contours and SMX and SMN shown on the top right hand side of the zoom-in views are their respective maximum and minimum values.

From Figs. 2.17a and 2.18a, we can see that the current density distributions after 10 ns and 100 ns are similar because the same current pulses are applied to the model. Higher current density (i.e., the red and orange regions in the zoom-in views of Fig. 2.17a) is found at the Via 12/Metal 1 interfaces due to current crowding. The location of the maximum current density depends on the direction of the current flow. For example, the highest current density at the PMOS region is found at the upper corners of the upper most contact at the PMOS source region as

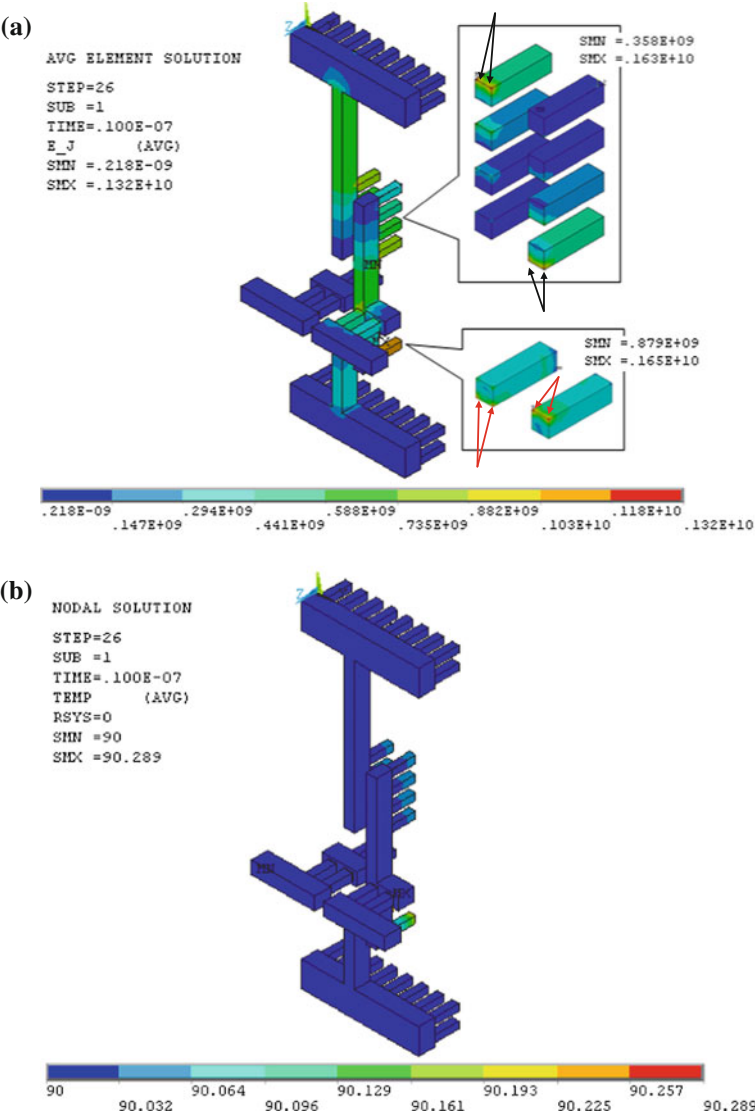


Fig. 2.17 Distributions of **a** current density (unit: $\text{pA}/\mu\text{m}^2$) and **b** temperature (unit: $^{\circ}\text{C}$) of the interconnects at the end of one load cycle (i.e. 10 ns)

the current is flowing from the V_{dd} line to the PMOS source, while it is found at the lower corners of the lowest contact at the PMOS drain region as the current is flowing from the PMOS drain to the output line, as indicated by the black arrows in Fig. 2.17a. For the NMOS contacts, the maximum current densities are at the lower corners of the source contact and the upper corners of the drain contact, as indicated by the red arrows.

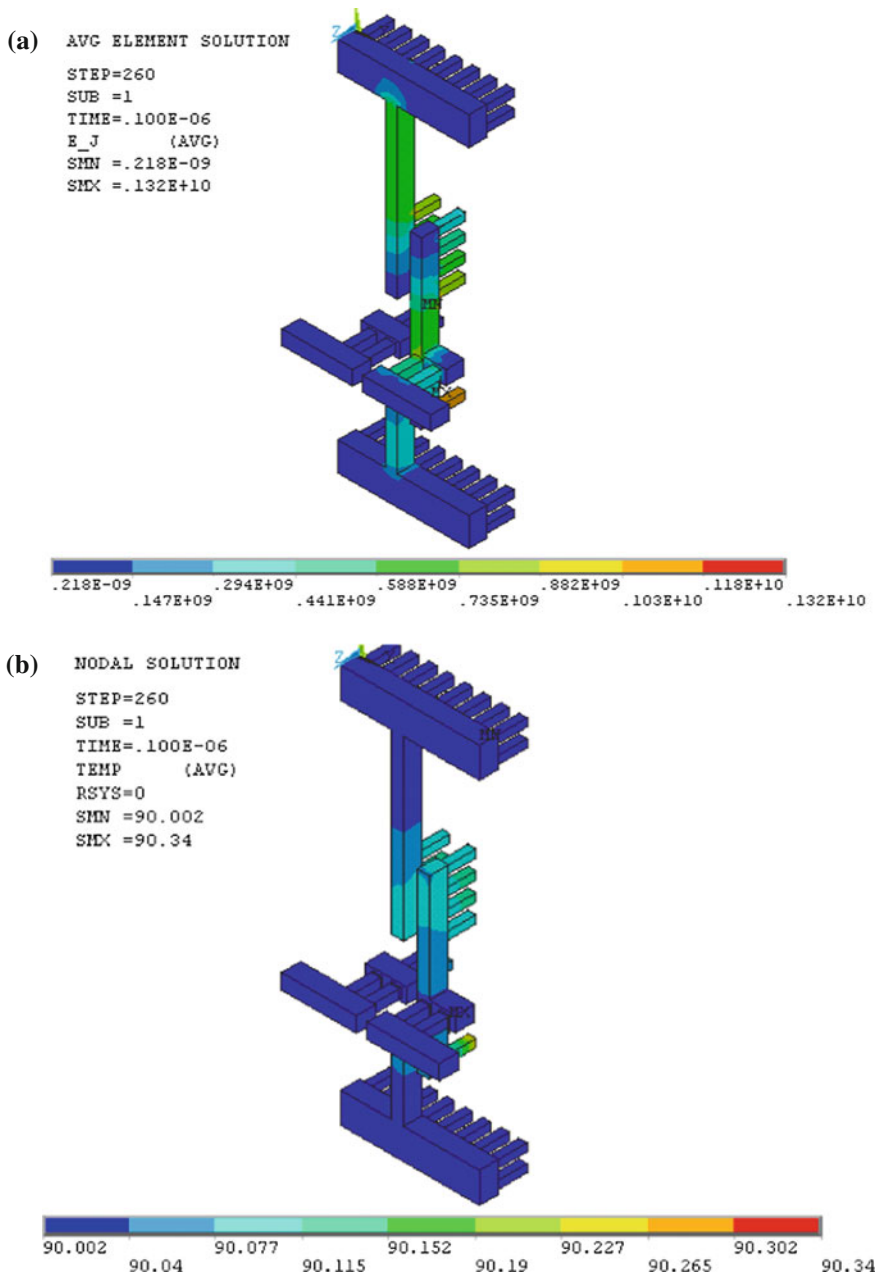


Fig. 2.18 Distributions of **a** current density (unit: $\text{pA}/\mu\text{m}^2$) and **b** temperature (unit: $^{\circ}\text{C}$) of the interconnects at the end of ten load cycles (i.e., 100 ns)

For the temperature distribution, higher temperature is found at the interconnect contact regions as they are nearer to the heat sources, i.e., the diffusion regions. This result is expected as the power consumption of the transistors is large due to the large load the transistors are driving. At the end of one load cycle, NMOS is on and PMOS is off, the temperature at the NMOS contacts at this instant is higher than that at the PMOS contacts (i.e., the light blue region for the PMOS contacts as compared with the green to red regions for the NMOS contacts as shown in Fig. 2.17b). The maximum temperature of the NMOS and PMOS contacts are 90.29 and 90.05 °C respectively. As time progresses, the temperatures at the diffusion regions increase, and this renders an increase in the interconnect temperature, gradually from the contact regions to the upper Metal 1, Via 12 and Metal 2 layers, as showing by the increase in area of the light blue regions in Fig. 2.18b. The location of the maximum temperature remains unchanged.

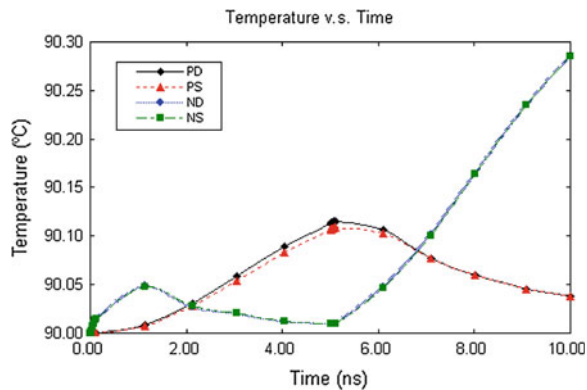
In the case of the inverter trees with hundreds or thousands of inverters working together, the circuit temperature is expected to be higher due to the interaction of the inverters with each other. The problem is even worse when the loads and the activities of the inverter trees are heavier (e.g., with more fan outs and/or at higher operating frequencies). The increase in temperature increases the total AFD in the interconnects.

2.4.2 Transient Temperature Variation During Circuit Operation

The transient temperature curve at various locations of the model can be plotted from the extracted results of the thermal-electric simulation. Figure 2.19 shows the time variation of the temperature change within one load cycle at the interconnect contact regions.

From Fig. 2.19, it is observed that the temperatures of the PMOS contacts gradually increase from time 0 to 5 ns (i.e., when V_{in} is at 0 V and PMOS is on).

Fig. 2.19 Temperature change with time at the interconnect contact regions within one load cycle



A slight increase in the temperatures of the NMOS contacts at the beginning is due to the current spike occurred during switching. The temperatures of the NMOS contacts drop slowly thereafter, but they do not drop to their initial values.

From 5 to 10 ns, NMOS turns on and PMOS turns off. The temperatures of the PMOS contacts gradually decrease due to heat loss to the surrounding, and they also do not fall back to their initial values. The temperatures of the NMOS contacts start to increase at 5 ns when NMOS begins to operate.

The increase in temperature for the NMOS contacts (around $0.29\text{ }^{\circ}\text{C}$) when NMOS is on is higher than that for the PMOS contacts (around $0.11\text{ }^{\circ}\text{C}$) when PMOS is on. This is expected because the size of PMOS in this design is around 2.56 times larger than that of NMOS so as to achieve an electrically symmetric characteristic (i.e., equal rise and fall time at the output), and thus the number of contacts that can be placed at the PMOS region (four pairs) is larger than that at the NMOS region (one pair). With the same current loads, the average current density of the NMOS contacts is higher than that of the PMOS contacts during their respective switch-on period, and thus a higher temperature rise for the NMOS contacts.

For a continuous circuit operation, the temperature variation at the interconnect contact regions shows a zig-zag shape corresponds to the turning on and off of the transistors. A general increasing trend in temperature is observed as shown in Fig. 2.20, which is also in agreement with the simulation result of Alam et al. [36].

Based on Fig. 2.20, the final temperature of the interconnect contact can be obtained through extrapolation using the curve fitting method. The extrapolated curve for the contact at NMOS drain is shown in Fig. 2.21. The final temperature is assumed to be the temperature at time tends to be infinity, which is $106.90\text{ }^{\circ}\text{C}$ in this case.

Fig. 2.20 Temperature change with time at the interconnect contact regions for multiple load cycles. A general increasing trend in temperature is observed

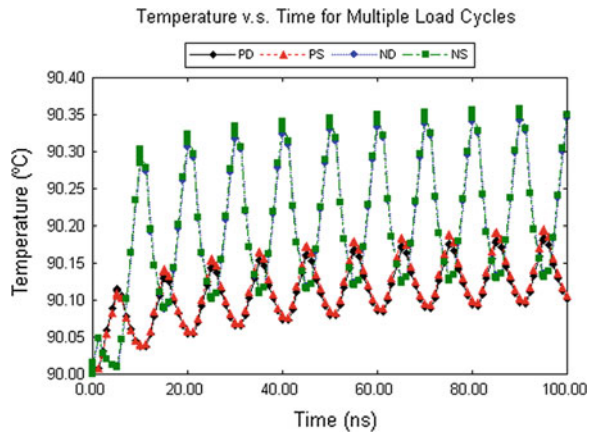
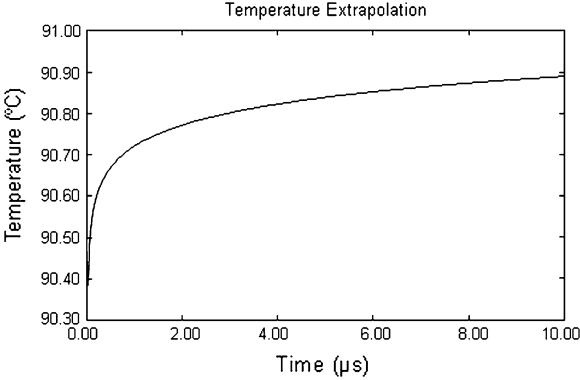


Fig. 2.21 Temperature extrapolation for the contact at NMOS drain

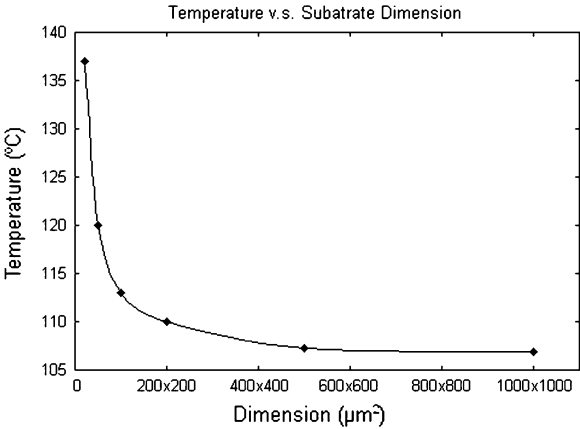


2.4.3 Effect of the Substrate Dimension on the Circuit Temperature

It is expected that the substrate dimension (length $\times$ width) can affect the heat dissipation and the circuit temperature. Keeping all other parameters the same, six models with different substrate areas varying from 20×20 to $1,000 \times 1,000 \mu\text{m}$ with the substrate thickness fixed at $300 \mu\text{m}$ are simulated under the same operation condition, and the results are shown in Fig. 2.22.

As the convection of the chip surface to the surrounding has minor impact on the circuit temperature as already explained in Sect. 2.3.1.1, a lower final steady-state temperature is observed for models with larger substrate areas due to their larger conduction areas at the heat sink (i.e., the metal base plate). The rise in the final circuit temperature is small when the substrate dimension decrease from $1,000 \times 1,000$ to $200 \times 200 \mu\text{m}$. A very rapid increase in temperature is observed when the dimension of the substrate area is smaller than $200 \times 200 \mu\text{m}$, showing

Fig. 2.22 Final circuit temperature of the inverter model with different substrate dimension. The substrate thicknesses of the models are fixed at $300 \mu\text{m}$



that the rate of heat loss to the surrounding through the small area is not enough to cater for the rate of increase in temperature of the circuit.

The above results further emphasizes the necessity of the complete 3D circuit model with the consideration of the size of the surrounding materials as in the actual physical implementation of the circuit.

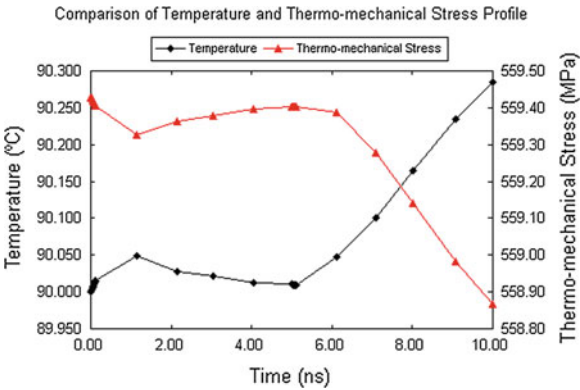
2.4.4 Transient Thermo-Mechanical Stress Variation During Circuit Operation

The time evolution of the thermo-mechanical stress at any location of the model can be obtained from the results of the transient structural-thermal simulation of the 3D model, and it is as shown in Fig. 2.23. The thermo-mechanical stress distribution of the interconnects at the end of one load cycle (10 ns) is shown in Fig. 2.24.

Intrinsic stress is present at the beginning of the simulation due to the thermal expansivity mismatch between the metallization and its surrounding materials when the metallization is cooled down from the process temperature (i.e. the SFT) to the circuit operation temperature. The total thermo-mechanical stress in the interconnects is the resultant stress of the intrinsic stress and the stress due to the temperature difference in the interconnects caused by the thermal loads. The magnitude of the total thermo-mechanical stress depends on the difference between the SFT of the metallization and the interconnect temperature.

When the interconnect temperature increases and approaches the SFT, the total thermo-mechanical stress in the interconnects decreases and shows an inverse relationship with the interconnect temperature as can be seen in Fig. 2.23. The change in the total thermo-mechanical stress at the end of one load cycle is only around 0.56 MPa and is much smaller than the intrinsic stress which is 559.49 MPa at 90 °C. The change in thermo-mechanical stress is small because the increase in temperature at the end of one load cycle is small (about 0.29 °C).

Fig. 2.23 Comparison of the temperature and thermo-mechanical stress profile of the interconnect contact at NMOS drain



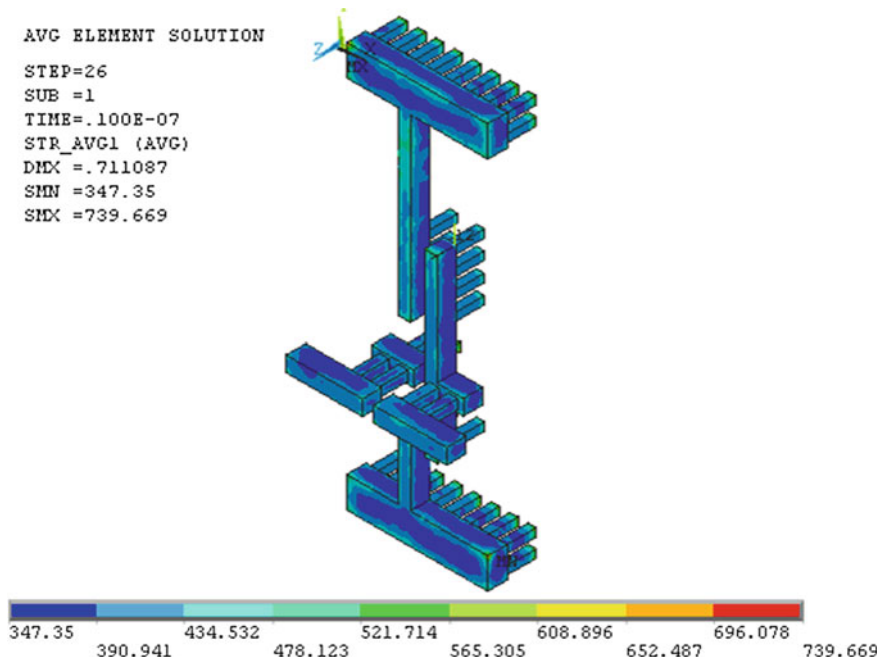


Fig. 2.24 Thermo-mechanical stress distribution (unit: MPa) of the interconnects at the end of one load cycle

Therefore the thermo-mechanical stress shown in Fig. 2.24 is mainly due to the intrinsic stress in the interconnects, which is more than 50 % higher at the corners (i.e., in contact with other materials) than in the center.

2.4.5 Distributions of the Atomic Flux Divergences

With the current density, temperature, temperature gradient and thermo-mechanical stress gradient obtained from the electro-thermo-structural simulations, the AFDs due to the three driving forces and the total AFD in the interconnects at a specific load step can be obtained.

Figure 2.25 shows the total AFD distribution of the interconnects at the end of one load cycle. The AFD distributions due to electron wind force, temperature gradient induced driving force and thermo-mechanical stress gradient induced driving force are shown in Fig. 2.26a, b, and c respectively.

With recent process improvement on the use of metal cap on Cu interconnects, one can assume a strong Cu/cap layer interface [37], and the void nucleation location (i.e. the EM weak spot) is found at the contact or via bottom as determined by the location of the maximum total AFD [38]. At the end of one load

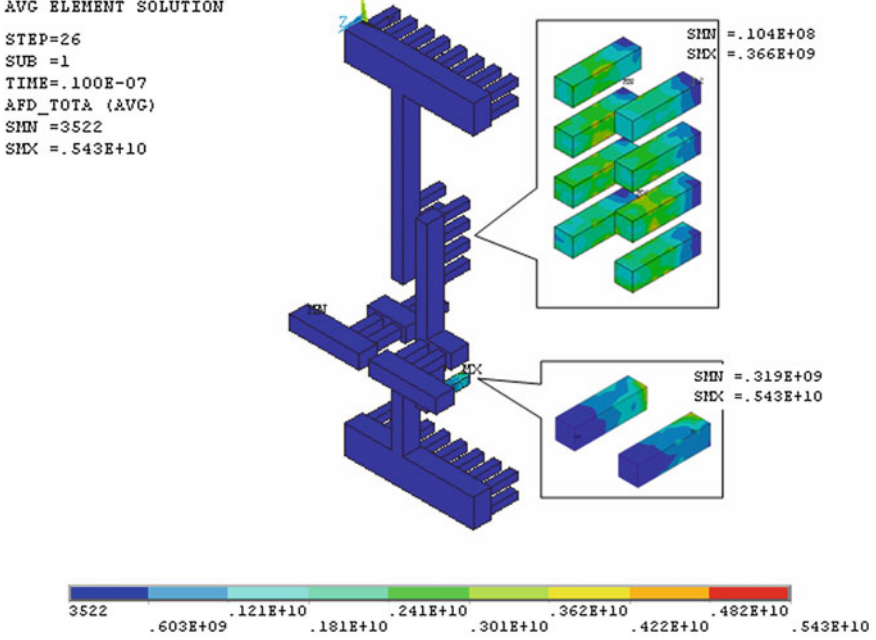


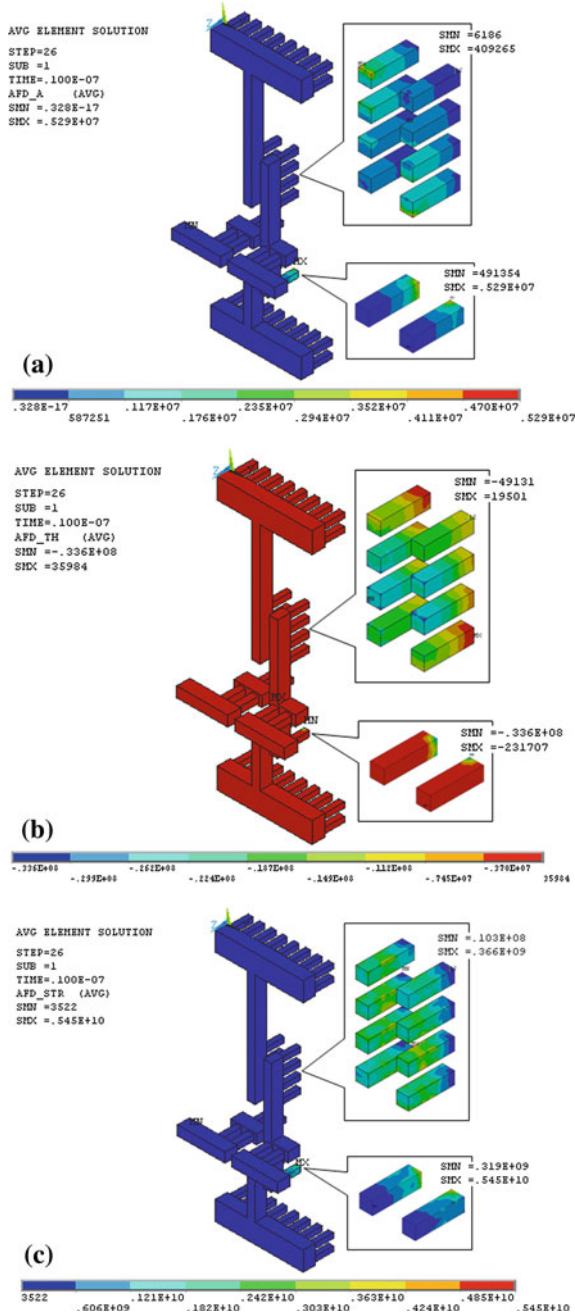
Fig. 2.25 Total AFD distribution (unit: atoms/ $\mu\text{m}^3\cdot\text{s}$) of the interconnects at the end of one load cycle. The figures on the right are the zoom-in views of the interconnect contacts at the PMOS and NMOS regions

cycle, the maximum AFD location of the inverter model is found at the bottom of the NMOS contacts as indicated by the arrows in Fig. 2.24, which is consistent with the work by Tan et al. [39] and Li et al. [35].

Comparing Figs. 2.25, 2.17a and 2.26a, it is clear that the total AFD distribution is very different from the distributions of the current density and the AFD due to electron wind force (AFD_EWF). At the end of one load cycle, AFD_EWF is 1 and 3 orders of magnitude smaller than the AFD due to temperature gradient induced driving force (AFD_T) and thermo-mechanical stress gradient induced driving force (AFD_S) respectively, as calculated from Eqs. (2.6)–(2.8). In fact, AFD_S has over 99 % contribution to the total AFD and the total AFD distribution of the model actually follows the distribution of AFD_S. In other words, if only current density is considered, the EM weak spot will be different from that if all the driving forces are considered.

From Eq. (2.6), we can see that AFD_EWF depends on the product of the current density and the temperature gradient of the interconnects. For the PMOS contacts, the locations of the maximum current density are the same as the maximum temperature gradient, as indicated by the black arrows in Fig. 2.17a and Fig. 2.27. The locations of the maximum AFD_EWF for the PMOS contacts are found at the upper corners of the upper most contact at the source region and the lower corners of the lowest contact at the drain region, as indicated by the black

Fig. 2.26 AFD distributions of the interconnects due to **a** electron wind force **b** temperature gradient induced driving force and **c** thermo-mechanical stress gradient induced driving force (unit: atoms/ $\mu\text{m}^3\cdot\text{s}$) at the end of one load cycle. The figures on the *right* are the zoom-in views of the interconnect contacts at the PMOS and NMOS regions



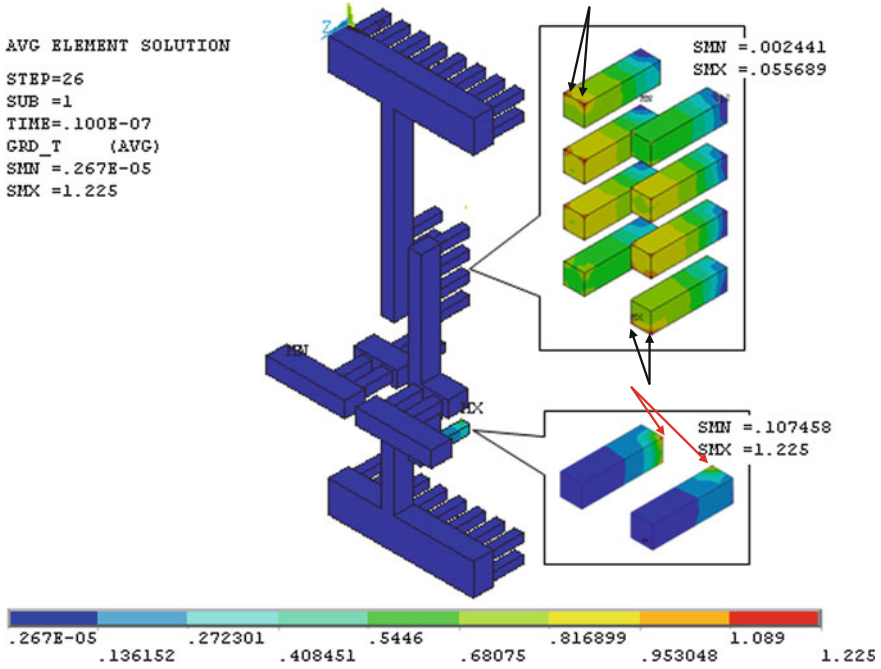


Fig. 2.27 Temperature gradient distribution of the interconnects (unit: $^{\circ}\text{C}/\mu\text{m}$) at the end of one load cycle. The figures on the *right* are the zoom-in views of the interconnect contacts at the PMOS and NMOS regions

arrows in Fig. 2.26a. For the NMOS contacts, the maximum current densities are at the top of the contacts as indicated by the red arrows in Fig. 2.17a, while the maximum temperature gradients are at the bottom of the contacts as indicated by the red arrows in Fig. 2.27. The temperature gradient at the bottom of the contacts is more than 10 times higher than that at the top while the current density at the top is only less than 2 times higher than that at the bottom. Therefore the locations of the maximum AFD_EWF are found at the bottom of the NMOS contacts, as indicated by the red arrows in Fig. 2.26a.

The maximum AFD due to AFD_T at the interconnect contact region is negative in value as shown in Fig. 2.26b. This is because at the locations with high temperature gradient and comparably small current density (e.g., the red to yellow regions in the zoom-in views of Fig. 2.27 and the blue regions in the zoom-in views of Fig. 2.17a), term 1 in Eq. (2.7) is negative due to its negative coefficient. It dominates over term 2 and results in a negative AFD_T (i.e., the blue to orange regions in the zoom-in views of Fig. 2.26b). At the region with low temperature gradient (e.g., the blue regions in the zoom-in views and blue regions in the global view of Fig. 2.27), term 2 dominates and a positive AFD_T is obtained.

From Eq. (2.8), term 1 dominates over term 2 and term 3 at 90°C and AFD_S depends on the product of the temperature gradient and the thermo-mechanical

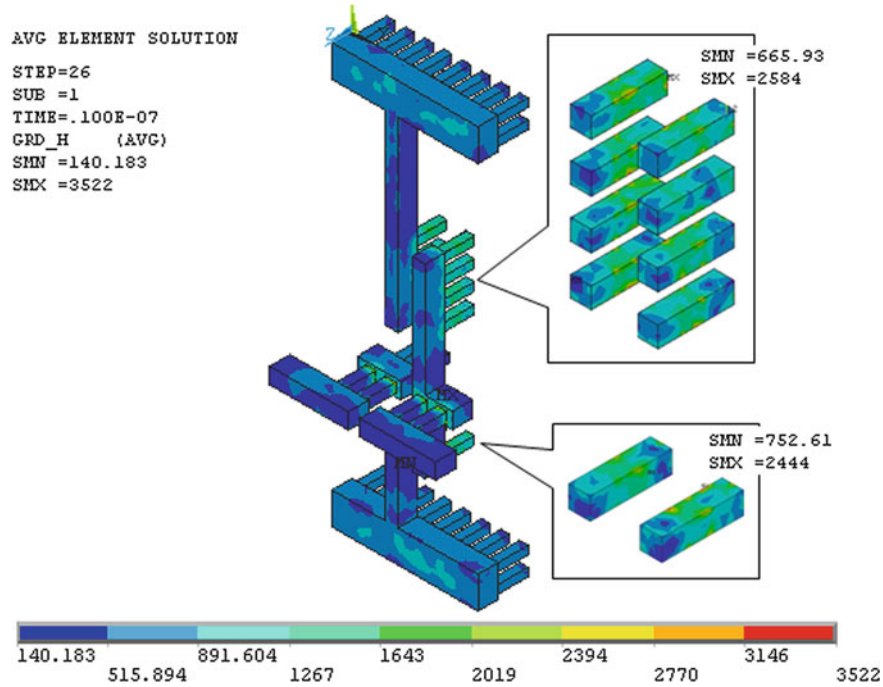


Fig. 2.28 Thermo-mechanical stress gradient distribution (unit: $\text{MPa}/\mu\text{m}$) of the interconnects at the end of one load cycle. The figures on the *right* are the zoom-in views of the interconnect contacts at the PMOS and NMOS regions

stress gradient. The thermo-mechanical stress gradient distribution of the interconnects is shown in Fig. 2.28. This stress gradient is mainly caused by the intrinsic stress in the interconnects. Similar distributions are found for the PMOS and NMOS contacts as they have similar intrinsic stresses. The values of the stress gradient at the two regions are also similar. At the regions where the temperature gradient are less than $0.06\text{ }^{\circ}\text{C}/\mu\text{m}$ (e.g. PMOS contacts, Metal 1, Via 12 and Metal 2), the AFD_S distribution follows the thermo-mechanical stress gradient distribution in the interconnect. At the regions with comparably large temperature gradient (e.g., the bottom of the NMOS contact), the AFD_S distribution follows the temperature gradient distribution. Therefore, to reduce AFD_S and hence the total AFD of the model, one can reduce the thermo-mechanical stress gradient and/or the temperature gradient. This will be studied in Sect. 4.4.

At the beginning of the circuit operation, the temperature and the resultant thermo-mechanical stress of the model, and hence the temperature gradient and the thermo-mechanical stress gradient are changing with time due to the variation in the electrical and thermal loads. The total AFD of the interconnects also changes with time as a result of the changing temperature gradient and thermo-mechanical stress gradient. However, as time progresses, the circuit temperature and temperature gradient become constant and are found to be the highest at the

bottom of the NMOS contacts (i.e., the same locations as in Figs. 2.17b and 2.27 but with a higher value). Therefore we can conclude that the maximum total AFD location at the end of one load cycle is the final maximum total AFD location of the model.

In summary, the simulation results in this section show the transient temperature response to the changes in electrical loads and the transient thermo-mechanical stress response to the changes in circuit temperature. This demonstrates the ability of the proposed 3D model for performing the transient temperature and stress analysis at circuit layout level. The distributions of the AFDs are discussed based on the current density, temperature gradient, and thermo-mechanical stress gradient distributions of the interconnects. The void nucleation location and hence the EM weak spot is found at the location with the highest total AFD, which is at the bottom of the NMOS contact for the inverter model under study. It is also found that the electron wind force is no longer the sole driving force for EM as the value and the distribution of the total AFD of the 3D circuit model are determined by AFD_S. To reduce AFD_S and hence the total AFD, the thermo-mechanical stress gradient and/or the temperature gradient of the circuit should be reduced.

2.5 Effects of Barrier Thickness and Low- κ Dielectric on Circuit EM Reliability

This section aims to show the structural and material modeling capability of the 3D circuit model by studying the effects of barrier layer thickness and low- κ dielectric on the EM reliability of the circuit.

Copper has a lower resistivity and a higher melting temperature than Aluminum, and it is replacing the traditional Al-alloy to reduce the RC delay of circuits. However, the high diffusivity of Cu in Si and SiO₂ demands a diffusion barrier layer around the Cu interconnects to prevent the out-diffusion of Cu [40]. A commonly used material for the barrier layer is Ta [41]. As the device feature size goes to 0.18 μm and below, thick barrier layer is found to be undesirable in circuit application due to its high total interconnect resistance [42]. On the other hand, the effectiveness of a diffusion barrier may degrade if its thickness is too thin. Lu et al. [43] showed that thin barrier rendered an early failure and resulted in a shorter EM lifetime due to Cu out-diffusion.

Also, low- κ dielectric is replacing the conventional SiO₂ ILD to further reduce the RC delay so as to meet the demand on the circuit switching speed. However, with a poorer interfacial adhesion and a smaller material elastic modulus, the reliability of the Cu/low- κ structure is a concern [44, 45]. Furthermore, the thermal conductivity of the low- κ dielectric is smaller than that of SiO₂, and this increases the circuit temperature and causes a faster degradation. Using the inverter circuit model in the previous section as an example, the final circuit temperature is 106.90

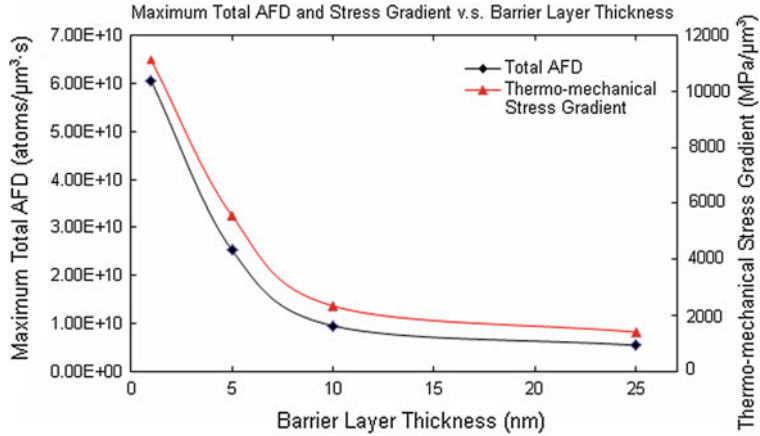


Fig. 2.29 Maximum total AFD and thermo-mechanical stress gradient versus barrier layer thickness for the contact at NMOS drain

and 133.02 °C for the SiO₂ and low- κ (carbon doped oxide) dielectric respectively. The temperature is much higher for the model using the low- κ dielectric due to its poorer thermal conductivity (0.58 W/m·°C) as compared with SiO₂ (1.38 W/m·°C).

To study the effect of barrier layer thickness on the EM reliability of the circuit, four models with the barrier layer thickness varying from 1 to 25 nm are simulated with SiO₂ ILD. The setup for the simulation are the same as in Sects. 2.2 and 2.3 and the simulation results at the end of one load cycle are shown in Fig. 2.29.

From Fig. 2.29, the maximum total AFD is found to increase with decreasing barrier layer thickness, and the rate increases by around 30 times when the barrier layer thickness goes below 7.5 nm, indicating a significant degradation in the EM lifetime with a barrier of thinner than 7.5 nm.

The percentage contribution of the three driving forces AFD_EWF, AFD_T, and AFD_S to the total AFD with different barrier layer thickness at the end of one load cycle is shown in Table 2.6. We can see that AFD_S dominates over the other two driving forces, especially for the models with thinner barriers.

As explained in the previous section, the value of AFD_S at 90 °C is determined by the product of the thermo-mechanical stress gradient and the temperature gradient. With the same electrical loads, the maximum temperature gradients of

Table 2.6 Percentage contribution of the three driving forces to the total AFD

Barrier thickness (nm)	Percentage contribution		
	AFD_EWF	AFD_S	AFD_T
1	0.017	99.855	0.128
5	0.051	99.632	0.317
10	0.075	99.393	0.532
25	0.097	99.284	0.619

the four models with different barrier layer thickness are similar, and it is the difference in the maximum thermo-mechanical stress gradients that causes the difference in the maximum AFD_S and total AFD.

As compared with Cu, Ta has a lower thermal expansion coefficient ($6.30 \times 10^{-6} \text{ }^{\circ}\text{C}^{-1}$ and $1.80 \times 10^{-5} \text{ }^{\circ}\text{C}^{-1}$ respectively) and a higher elastic modulus ($1.86 \times 10^5 \text{ MPa}$ and $1.10 \times 10^5 \text{ MPa}$ respectively), the presence of the Ta barrier introduces a tensile stress in the Cu interconnects in the direction normal to the surface [28, 46]. When the barrier thickness decreases, the confinement effect from the barrier is reduced and this increases the thermo-mechanical stress gradient, as can be seen in in Fig. 2.29. The increase in the thermo-mechanical stress gradient in turn causes an increase in the maximum total AFD. A higher maximum AFD renders a shorter EM lifetime [47], therefore a thinner barrier can cause the reliability degradation of the circuit, and the effect is increasingly significant when the barrier layer thickness goes below 7.5 nm.

Keeping all other conditions the same, various low- κ dielectric, e.g. SiLK, MSQ (methyl silsesquioxane), CDO (carbon doped oxide), and SiCOH are used to replace the traditional SiO_2 ILD. The material properties of the low- κ dielectric are listed in Table 2.7, and the values of the maximum total AFD and thermo-mechanical stress gradient of the models with different low- κ dielectric are shown in Fig. 2.30. The barrier layer thicknesses of the models are kept at 25 nm.

As shown in Fig. 2.30, the models using the low- κ dielectric have a higher maximum total AFD than that using the SiO_2 ILD due to their higher thermo-mechanical stress gradients. This is because the low- κ materials have lower elastic modulus than SiO_2 , as shown in Table 2.7. Therefore the confinement effect provided by the elastic property of the low- κ material is lower as proven by the experimental work of Webb et al. [50]. Looking at the high stress gradients due to MSQ and CDO in Fig. 2.30, and that they have much lower elastic modulus than SiO_2 , one can expect that Cu with MSQ or CDO dielectric suffers significant reduction in the EM lifetime. In addition, as can be seen in Table 2.7, the low- κ materials have higher thermal expansion coefficient (CTE) than SiO_2 , therefore the difference in CTE between Cu and the low- κ materials is smaller than that between Cu and SiO_2 . This further reduces the confinement effect and increases the thermo-mechanical stress gradient [28, 46].

Table 2.7 The material properties of the low- κ dielectric as compared with SiO_2 [48, 49]

Materials	Properties				
	Young's modulus (MPa)	Poisson's ratio	Density (kg/m^3)	Thermal expansion ($1/^{\circ}\text{C}$)	Thermal conductivity ($\text{W/m}\cdot^{\circ}\text{C}$)
SiLK	2.50×10^3	0.40	900	6.60×10^{-5}	0.14
MSQ	3.60×10^3	0.25	890	7.30×10^{-6}	0.40
CDO	4.47×10^3	0.30	1,510	1.29×10^{-5}	0.58
SiCOH	1.62×10^4	0.30	910	1.20×10^{-5}	0.40
SiO_2	7.14×10^4	0.16	2,200	6.80×10^{-7}	1.38

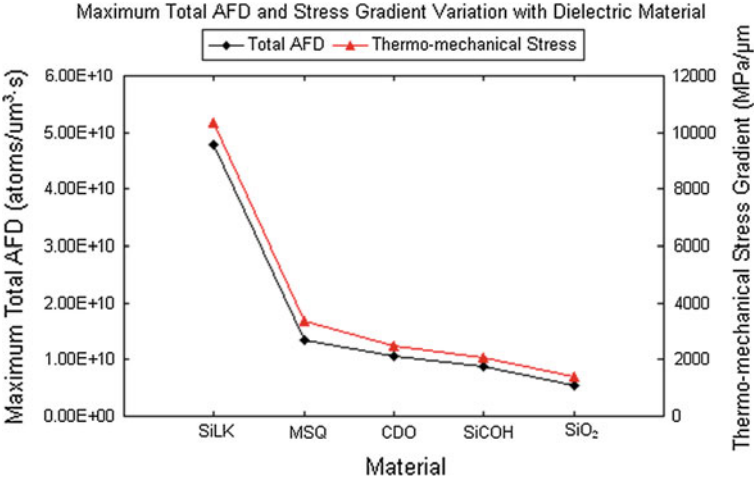


Fig. 2.30 The maximum total AFD and thermo-mechanical stress gradient variation with different dielectric materials for the contact at NMOS drain

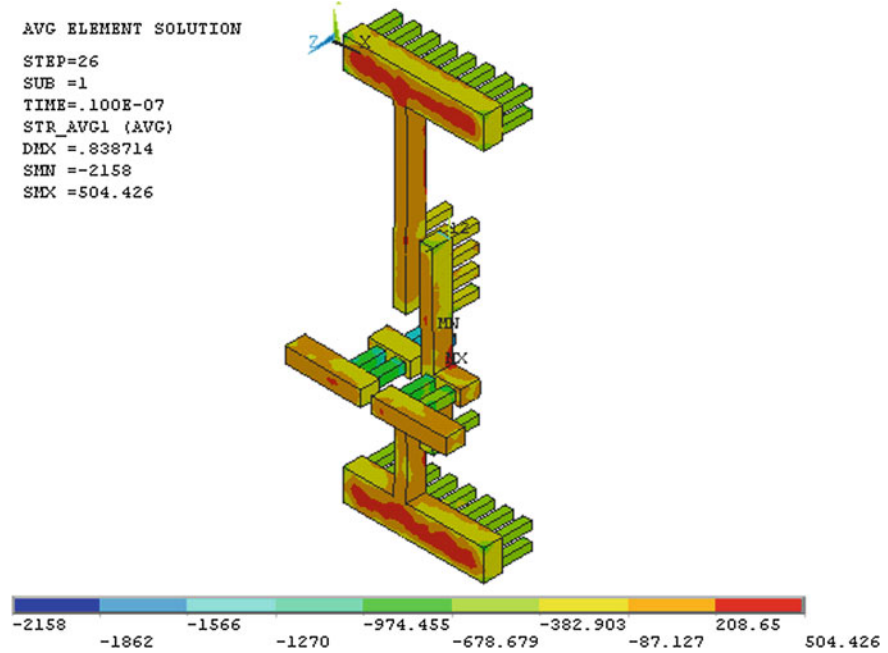


Fig. 2.31 Thermo-mechanical stress distribution in the Cu/SiLK structure (unit: MPa)

With varying line width on different parts of the interconnect structure, the relative barrier thickness and hence the structural confinement from the barrier and the dielectric are different [28]. In the Cu/SiLK structure, due to the large difference in CTE between Cu and SiLK, and that the CTE of SiLK is larger than Cu, the confinement effect provided by the Ta barrier at the regions with smaller interconnect dimension (e.g., the contact or via regions) is no longer enough to produce the positive tensile stress. As a result, a negative compressive stress is produced, as can be seen from the blue and green regions in Fig. 2.31. This results in a combination of the positive tensile and negative compressive stresses on different parts of the interconnects and thus increases the stress gradient of the Cu/SiLK structure. The stress state of this interconnect structure becomes dominated by shear stresses and the reliability concern changes to interfacial delamination [46].

In summary, this section presents the increase in the total AFD with decreasing barrier layer thickness and the use of the low- κ materials, using the inverter circuit model as an example. The increase in the total AFD with decreasing barrier layer thickness is due to the increase in the thermo-mechanical stress gradient with a thinner barrier. The low- κ dielectric with a lower elastic modulus and a higher CTE as compared with SiO₂ reduces the structural confinement and renders a higher thermo-mechanical stress gradient, and thus a higher total AFD.

2.6 Summary

This chapter demonstrated the step by step construction of a complete 3D circuit model from its 2D IC layouts. For ease of demonstration, a simple inverter circuit with only the intra-block interconnects was used as an example. The method of conducting the transient electro-thermo-structural simulations using both Cadence (a circuit simulator) and ANSYS (a finite element tool) was also introduced.

The current density, temperature and thermo-mechanical stress, as well as the atomic flux divergences (AFDs) at any load step and any node of the model could be determined and plotted from the simulation results, and hence the EM weak spots of the interconnects of the circuit, which is the location with the highest total AFD could be identified.

Different from other 3D EM models in literature, the 3D model presented here was able to perform the transient analysis at circuit layout level. The simulation results successfully showed the transient temperature and stress responses to the changes in electrical loads and circuit temperature respectively. Therefore the model was capable to evaluate the EM reliability of the interconnects of the circuit with different loadings and under different operation conditions.

To demonstrate the structural and material modeling capability of the 3D circuit model, the effects of barrier layer thickness and low- κ dielectric on the EM reliability of the circuit were studied using the inverter circuit model. It was found that, although the use of the sub-5 nm barrier layers in combination with a low- κ

dielectric was beneficial for reducing the RC delay, a higher maximum total AFD was observed in the structures with a thinner barrier or with a low- κ dielectric. A shorter EM lifetime was expected when the barrier thickness scaled below 7.5 nm and/or when the SiO₂ ILD was replaced by a low- κ dielectric.

References

1. Complete timing signoff in the nanometer era, http://w2.cadence.com/whitepapers/timing_signoff_wp.pdf
2. Baker RJ (2004) CMOS circuit design—layout and simulation, revised 2nd edn. IEEE Press Series on microelectronic Systems, Wiley, London
3. Tan CM, He F (2009) 3D circuit model for 3D IC reliability study. In: 10th International conference on thermal, mechanical and multi-physics simulation and experiments in microelectronics and microsystems (EuroSimE), 2009, p 7
4. Tan CM, Hou Y, Li W (2007) Revisit to the finite element modeling of electromigration for narrow interconnects. *J Appl Phys* 102(2):033705-1-7
5. He F, Tan CM (2010) Circuit level interconnect reliability study using 3D circuit model. *Microelectron Reliab* 50(3):376–390
6. He F, Tan CM (2010) Modeling the effect of barrier thickness and low- κ dielectric on circuit EM reliability using 3D model. *Microelectron Reliab* 50(9–11):1327–1331
7. Alvino WM (1994) *Plastics for electronics, materials, properties, and design applications*. McGraw-Hill, New York, p 326
8. Chartered Semiconductor Manufacturing. Document 0.18um DROIT_06 June 07
9. Chung DDL (1995) *Materials for electronic packaging*. Butterworth-Heinemann, London, p 57
10. Pecht M (1999) *Electronic packaging materials and their properties*. CRC Press, Boca Raton, p 151
11. Dalleau D, Weide-Zaage K, Danto Y (2003) Simulation of time depending void formation in copper, aluminum and tungsten plugged via structures. *Microelectron Reliab* 43(9–11):1821–1826
12. Chen CT, Hsu T-S, Jeng R-J, Yeh H-C (2000) Enhancing the glass-transition temperature of polyimide copolymers containing 2,2-bipyridine units by the coordination of nickel malenonitriledithiolate. *J Polym Sci Part A: Polym Chem* 38(3):498–503
13. Dalleau D, Weide-Zaage K (2001) Three-dimensional voids simulation in chip metallization structures: a contribution to reliability evaluation. *Microelectron Reliab* 41(9–10):1625–1630
14. Ciptokusumo J, Weide-Zaage K, Aubel O (2009) Investigation of stress distribution in via bottom of Cu-via structures with different via form by means of submodeling. *Microelectron Reliab* 49(9–11):1090–1095
15. Weide-Zaage K, Dalleau D, Danto Y, Fremont H (2007) Dynamic void formation in a DD-copper-structure with different metallization geometry. *Microelectron Reliab* 47(2–3):319–325
16. Gonzalez JL, Rubio A (1997) Shape effect on electromigration in VLSI interconnects. *Microelectron Reliab* 37(7):1073–1078
17. What temperature should my processor be running at, <http://www.computerhope.com/issues/ch000687.htm>
18. Chen D, Li E, Rosenbaum E, Kang S-M (2000) Interconnect thermal modeling for accurate simulation of circuit timing and reliability. *IEEE Trans Comp-Aided Des Integr Circuits Syst* 19:197–250

19. Lasance C, Moffat C (2005) Advances in high-performance cooling for electronics. *Electron Cooling* 11(4), <http://www.electronics-cooling.com/2005/11/advances-in-high-performance-cooling-for-electronics/>
20. Cheng Y-K, Tsai C-H, Teng C-C, Kang S-M (2002) *Electrothermal analysis of VLSI systems*. Kluwer Academic Publishers, Dordrecht
21. Rzepka S, Banerjee K, Meusel E, Hu CM (1998) Characterization of self-heating in advanced VLSI interconnect lines based on thermal finite element simulation. *IEEE Trans Compon Packag Manuf Tech Part A* 21(3):406–411
22. Lloyd JR (1999) Electromigration in integrated circuit conductors. *J Appl Phys* 32(17):R109–R118
23. Schoen JM (1980) A model of electromigration failure under pulsed condition. *J Appl Phys* 51:508
24. Tao J, Cheung NW, Hu C (1994) Electromigration failure model for interconnects under pulsed and bidirectional current stressing. *IEEE Trans Electron Devices* 41(4):539–545
25. Adler V, Friedman EG (1996) Delay and power expressions for a CMOS inverter driving a resistive-capacitive load. In: *IEEE international symposium on circuits and systems*, vol 4, pp 101–104
26. Chen WY (2004) *Home networking basics*. Prentice Hall, Englewood Cliffs
27. Shen Y-L, Ramamurty U (2003) Temperature-dependent inelastic response of passivated copper films: experiments, analyses, and implications. *J Vac Sci Tech B: Microelectron Nanometer Struct* 21(4):1258–1264
28. Rhee S-H, Du Y, Ho PS (2003) Thermal stress characteristics of Cu/oxide and Cu/low- κ submicron interconnect structure. *J Appl Phys* 93(7):3926–3933
29. Love AE (1927) *A treatise on the mathematical theory of elasticity*. Cambridge University Press, Cambridge
30. Released 11.0 Documentation for ANSYS, Chapter 9. Submodeling
31. Wang E, Nelson T, Rauch R (2004) Back to elements—tetrahedra vs. hexahedra. In: *International ANSYS Conference*, 2004
32. Wang TH (2005) Submodeling analysis for path-dependent thermomechanical problems. *Trans ASME J Electron Packag* 127(2):135–140
33. Christou A (1993) *Electromigration and electronic device degradation*. John Wiley, New York
34. Roy A, Tan CM (2006) Experimental investigation on the impact of stress free temperature on the electromigration performance of copper dual damascene submicron interconnect. *Microelectron Reliab* 46(9–11):1652–1656
35. Li W, Tan CM, Hou Y (2007) Dynamic simulation of electromigration in polycrystalline interconnect thin film using combined Monte Carlo algorithm and finite element modeling. *J Appl Phys* 101(10):104314
36. Alam SM, Gan CL, Thompson CV, Troxel DE (2007) Reliability computer-aided design tool for full-chip electromigration analysis and comparison with different interconnect metallizations. *Microelectron J* 38(4–5): 463–473
37. Petrov N, Valverde C (2005) Process control and material properties of thin electroless Co-based capping layers for copper interconnects. In: *Proceeding of SPIE*, 2005, vol 6002, pp 01–011
38. Sukharev V (2005) Physically based simulation of electromigration-induced degradation mechanisms in dual-inlaid copper interconnects. *IEEE Trans Comp-Aided Des Integr Circuits Syst* 24(9):1326–1335
39. Tan CM, Roy A (2006) Investigation of the effect of temperature and stress gradients on accelerated EM test for Cu narrow interconnects. *Thin Solid Films* 504(1–2):288–293
40. Attardo MJ, Rosenberg R (1970) Electromigration damage in aluminum film conductors. *J Appl Phys* 41:2381–2386
41. Wang MT, Lin YC, Chen MC (1998) Barrier properties of very thin Ta and TaN layers against copper diffusion. *Electrochem Soc* 145:2538–2545

42. Ou KL, Wu WF, Chiou SY (2007) Enhancing the reliability of $n^+ - p$ junction diodes using plasma treated tantalum barrier film. *Microelectron Eng* 84:151–160
43. Lu X, Pyun JW, Li B, Henis N, Neuman K, Pfeifer K, Ho PS (2005) Barrier layer effects on electromigration reliability of Cu/low k interconnects. In: *IEEE Proceedings of inter interconnect technology conference 2005*, pp 33–35
44. Paik JM, Park H, Joo YC (2004) Effect of low- κ dielectric on stress and stress-induced damage in Cu interconnects. *Microelectron Eng* 71:348–357
45. Cheng YL, Wang YL, Chen HC, Lin JH (2006) Effect of inter-level dielectrics on electromigration in damascene copper interconnect. *Thin Solid Films* 494:315–319
46. Zhao J-H, Qi W-H, Ho PS (2002) Thermomechanical property of diffusion barrier layer and its effect on the stress characteristics of copper submicron interconnect structures. *Microelectron Reliab* 42:27–34
47. Tan CM (2010) *Electromigration in ULSI interconnects*. World Scientific, Singapore
48. Lee HJ, Christopher SL, Liu DW, Bauer BJ, Lin EK, Wu WL, Grill A (2004) Structural characterization of porous low- κ thin films prepared by different techniques using x-ray porosimetry. *J Appl Phys* 95:2355–2359
49. Chen F, Gill J, Harmon D, Sullivan T, Li B, Strong A, Rathore H, Edelstein D, Yang CC, Cowley A, Clevenger L (2004) Measurements of effective thermal conductivity for advanced interconnect structures with various composite low- κ dielectrics. In: *IEEE International Reliability Physics Symposium*, 2004, pp 68–73
50. Webb E, Witt C, Andryuschenko T, Reid J (2004) Integration of thin electroless copper films in copper interconnect metallization. *J Appl Electrochem* 34:291–300

Electromigration Modeling at Circuit Layout Level

Tan, C.M.; He, F.

2013, IX, 103 p. 75 illus., 2 illus. in color., Softcover

ISBN: 978-981-4451-20-8