PAPER

Prediction of Self-Heating in Short Intra-Block Wires

Ken-ichi SHINKAI[†a)], ***Student Member***, **Masanori HASHIMOTO**[†b)], ***and*** **Takao ONOYE**[†c)], ***Members***

SUMMARY This paper investigates whether the self-heating effect in short intra-block wires will become apparent with technology scaling. These wires seem to have good thermal radiation characteristics, but we validate that the self-heating effect in local signal wires will be greater than that in optimal repeater-inserted global wires. Our numerical experiment shows that the maximum temperature increase from the silicon junction temperature will reach 40.4°C in a steady state at a 14-nm process. Our attribution analysis also demonstrates that miniaturizing the area of wire cross-section exacerbates self-heating as well as using low-κ material and increased power dissipation in advanced technologies below 28 nm. It is revealed that the impact of self-heating on performance in local wires is limited, while underestimating the temperature may cause an unexpected reliability failure.

key words: *self-heating, temperature, interconnect, local wire, process scaling*

1. Introduction

Thermal integrity has recently been recognized as an important design topic, because leakage current and reliability such as electromigration [1], [2] strongly depend on temperature, while increased power consumption causes the temperature rise. Therefore, the thermal analysis of a whole chip and temperature-aware design from the architectural level [3] to the physical design level [4] are hot topics in both academia and industry.

From a microscopic point of view, the self-heating problem in global wires has been studied intensively [5]–[11]. The self-heating effect is a phenomenon in which the energy dissipated in a wire resistance generates the heat inside a wire, and hence results in the temperature increase. An overheated wire degrades signal propagation (due to the increase in electrical resistance) and reliability [8]. The self-heating problem in global wires is thought to be significant for the following reasons [8]: (1) global wires are located far from a silicon (Si) substrate and thus the generated heat is unlikely to diffuse, and (2) a large current that usually flows in a global wire generates a lot of heat inside it.

In contrast, local and intermediate wires have been paid less attention, because their temperature increases do not seem to be significant due to their adjacency to the substrate, which has excellent thermal conductivity. However, process scaling raises three design problems that may aggravate self-heating. First, it significantly elevates the local clock frequency, which is higher than the global clock frequency. Second, it increases thermal resistance as a result of shrinking the wire cross-sectional area. Third, miniaturized devices use low-κ materials as insulators. Hence, it is not clear whether the self-heating effect in short intra-block wires will be a problem in the future.

In this paper, we evaluate the self-heating effects of local signal, global signal and local power wires. Our numerical experiments show that the temperature difference between local power lines and Si substrate is less than 4.74°C, although the temperature difference between local signal wires and substrate can reach 40.4°C. In addition, we evaluate degradation of reliability and performance caused by self-heating in local wires. It is found that reliability could be exacerbated, while the performance degradation will be relatively small despite the large increase in temperature.

A preliminary evaluation of self-heating in local signal wires was reported in [12]. In this work, we reexamine the prediction setup considering the size effect of Cu interconnects and the mutual dependence between temperature and wire resistances. We have found that the size effect increases the self-heating of local wires, but it makes the wire delay less dependent on temperature. We also investigate the accuracy of thermal analysis based on a finite-difference approach in terms of the mesh fineness, and reduce the estimation error using an extrapolation. Our attribute analysis with respect to both local and global signal wires clearly demonstrates that the smaller cross-sectional area is a remarkable factor that increases temperature of local wires, and it is less influential in global wires. Moreover, local power wires are studied in order to make the analysis of local warming more comprehensive.

The rest of the paper is organized as follows. Section 2 explains the motivation of this work. Section 3 describes an analysis technique, device parameters, and models used for prediction. Section 4 presents our experimental results for the increase in temperature and demonstrates an attribution analysis that examines which scaling parameter greatly influences temperature. We also evaluate the degradation of reliability and performance. Finally, we conclude the paper in Sect. 5.

2. Motivation

In this section, to explain why we focus on the self-heating

Manuscript received June 10, 2009.
Manuscript revised September 16, 2009.

[†]The authors are with the Department of Information Systems Engineering, Osaka University, Suita-shi, 565-0871 Japan.
a) E-mail: shinkai.kenichi@ist.osaka-u.ac.jp
b) E-mail: hasimoto@ist.osaka-u.ac.jp
c) E-mail: onoye@ist.osaka-u.ac.jp

DOI: 10.1587/transfun.E93.A.583

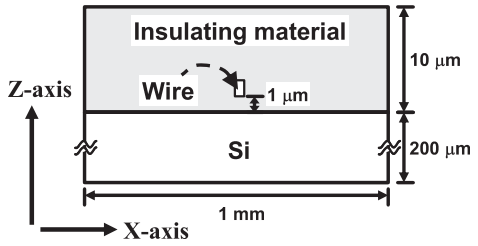

Fig. 1 2D chip model.

Table 1 Parameter set for 2D chip model.

Parameter		Year (Tech. node [nm])					
		2005 (90)	2008 (59)	2011 (40)	2014 (28)	2017 (20)	2020 (14)
Aspect ratio of wire		1.7	1.8	1.9	1.9	2.0	2.0
Power consumption in wire [μW/μm]		0.034					
Metal insulator (Thermal conductivity [W/m·K])		Fluorinated silicate glass (FSG) 0.89					
Heat transfer coefficient [W/m^{2}K]	Top	2500					
	Bottom	7500					
	Lateral sides	1 (insulated)					

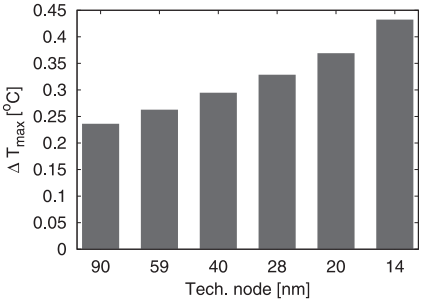

Fig. 2 Prediction of ΔT_{max} (2D chip model).

problem in short intra-block wires, we demonstrate a simple example of how shrinking the cross-sectional area aggravates it. Let us assume the simplified 2D chip model shown in Fig. 1. The approach to predict self-heating is identical to one which will be described in Sect. 3.1 except that here it uses 2D model. The center rectangle represents a Cu wire, and the lower rectangle represents a Si substrate. The parameter set used for this simulation is shown in Table 1. Aspect ratio of wire and power consumption are based on ITRS 2005 prediction [13] and others are determined similarly to Sect. 3.2.2. The power consumption of 0.034 μW/μm corresponds to the value calculated with J_{max}, the conductor effective resistivity and the local wire cross-sectional area at 90-nm process. To see the effect of the scaled wire cross-sectional area clearly, in this example, only wire width and aspect ratio of wire are assumed to scale according to process technology.

Figure 2 shows the increase in temperature of local wires in future processes. ΔT here is the temperature increase from the temperature of the Si junction, and ΔT_{max} is the maximum of ΔT. We can see that the increase in temperature is caused by the miniaturization of the wire cross-sectional area, because the other parameters are constant.

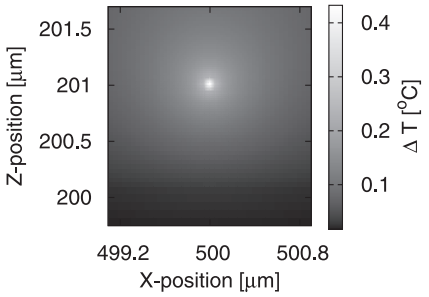

Fig. 3 Distribution of ΔT in 2020 (2D chip model). Axes are indicated in Fig. 1.

The distribution of the temperature increase from the substrate temperature in 2020 (Tech. node: 14 nm) is depicted in Fig. 3. It illustrates that a smaller cross-section impedes the heat dispersion, and thus the temperature rise is caused in the area immediately around the wire. Despite this finding, the degradation of thermal conductivity caused by the miniaturization of the wire cross-sectional area has not been regarded as a major problem. This fact motivates us to study self-heating in local wires.

3. Setup for Prediction

Experiments in this paper are performed with local/global signal wires and local power wires. This section shows how to perform analyses, and describes structure models and device parameters used to predict self-heating.

3.1 Analysis Flow

In a thermal analysis, we use a finite-difference approach [4], [14] to solve the heat diffusion equation based on the electrical-thermal analogy [7]. First we simulate an actual electrical circuit, and then a thermally-equivalent circuit using analyzed information about power consumption. To take the mutual dependence between temperature and wire resistances into consideration, we iteratively compute the heat generated in a wire and the temperature with the flowchart shown in Fig. 4 and obtain a self-consistent steady-state temperature. A convergence check is performed by comparing the increase in temperature with that in the previous loop. When the difference between those becomes small enough, iterations are stopped. In our analysis, two iterations are appropriate. The following subsections will describe each step.

3.1.1 Electrical Circuit Simulation

For 3D thermal simulation, power consumption is calculated with transistor-level circuit simulation using HSPICE. This electrical circuit simulation aims to take into account the nonuniform heating in a wire shown in Fig. 5. When a CMOS gate is switching, a larger current flows and hence higher temperatures are generated in the wire near the driver side than the receiver side.

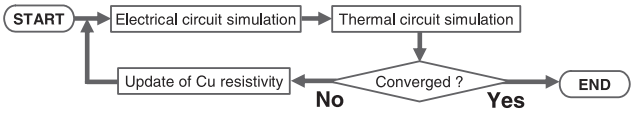

Fig. 4 Analysis flowchart for obtaining steady-state temperature in a self-consistent manner.

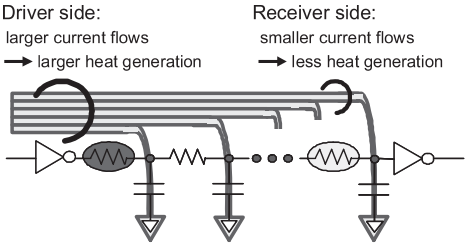

Fig. 5 Nonuniform heating in a wire.

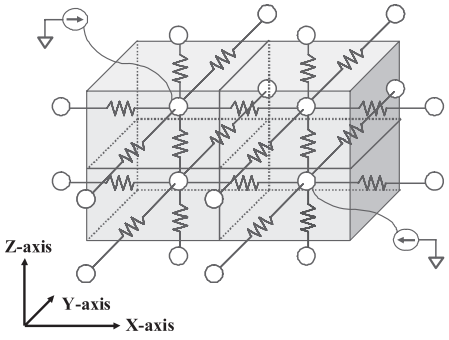

Fig. 6 An example of heat diffusion model.

We measure the current flow through each RC section and obtain power consumption of RI^2 for each section. After that, the average power consumption is calculated, because a steady state is analyzed in this work. A section corresponds to a grid used for the thermal circuit simulation.

3.1.2 Thermal Circuit Simulation

In a finite-difference approach [4], [14], a chip is first divided into cuboids. Next, thermal resistances are connected between adjacent nodes, where nodes are placed at the center of each cuboid. Finally, every consumed power calculated in Sect. 3.1.1 is injected as a current from ground into the corresponding node. The temperature in a steady state is calculated by DC analysis with a circuit simulator. As a simple example, when a chip is evenly divided in lengthwise and crosswise directions, the model becomes like Fig. 6. Our thermal analysis is conducted with a high-speed linear circuit simulator [15], [16] on a Linux server with AMD Opteron 2.8 GHz and 16 GB memory.

We here examine the setup of thermal analysis in terms of discretization. A previous work [7] reported that experimental results obtained with HSPICE-based thermal circuit simulation and finite element thermal simulation, e.g. ANSYS [17], are consistent, if discretization is performed appropriately. When a chip is finely discretized, the error becomes small, while more time and memory are necessary for thermal analysis. We therefore mitigate the temperature estimation error due to discretization with an extrapolation. To evaluate the error originating from discretization, we first execute thermal analysis with various discretization conditions. Before discussing it, let us define discretization of $54{\times}a$ in x, $37{\times}b$ in y and $23{\times}c$ in z as $(a, b, c)^{\dagger}$. Figure 7 shows that the relationship between the accuracy and computational time for the case that the number of sections in a single direction increases. As an example of changing the number of sections, Fig. 8 demonstrates a case that the number of partitions in x direction is changed from 1 to 2. Here, ΔT_{max} as an index of accuracy is evaluated similarly to that in Sect. 4. We can see that as the number of discretization increases, the result converges but the computational time increases.

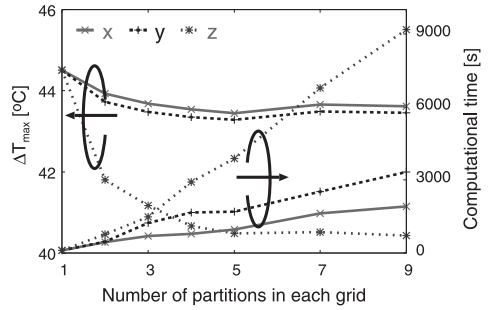

Fig. 7 Trade-off between accuracy and computational time.

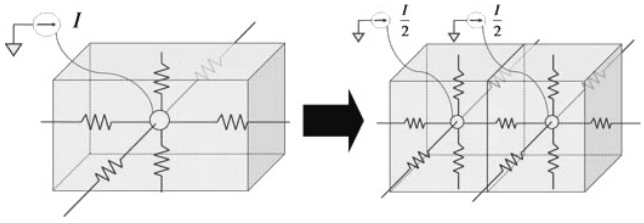

Fig. 8 A mesh changed from (1, 1, 1) to (2, 1, 1).

The correct result would be estimated with (∞, ∞, ∞) but it is impossible to compute it directly. On the other hand, we approximate the temperature with (a, b, c) as Eq. (1) using the first-order Taylor expansion.

$$T(a,b,c) \simeq T(1,1,1) + \frac{\partial T}{\partial a}\Delta a + \frac{\partial T}{\partial b}\Delta b + \frac{\partial T}{\partial c}\Delta c, \quad (1)$$

where T is the temperature at a node. Δa, Δb and Δc are the difference of a, b and c from the value of 1. $\frac{\partial T}{\partial a}$ is the partial derivative of T with respect to a at (a, b, c)=(1, 1, 1). $\frac{\partial T}{\partial b}$ and $\frac{\partial T}{\partial c}$ are similar to $\frac{\partial T}{\partial a}$. When we compute $\frac{\partial T}{\partial a}$ as $\frac{T(a,1,1)-T(1,1,1)}{a-1}$, and $\frac{\partial T}{\partial b}$ and $\frac{\partial T}{\partial c}$ similarly, Eq. (1) is expressed as Eq. (2).

$$T(a,b,c) \simeq T(1,1,1) + \Big\{T(a,1,1) - T(1,1,1)\Big\}$$

†To accurately analyze temperature, we divide a chip into fine sections near wires in particular. The minimal mesh size with (1, 1, 1) is equal to the wire width.

$$+\Big\{T(1,b,1)-T(1,1,1)\Big\}+\Big\{T(1,1,c)-T(1,1,1)\Big\}. \quad (2)$$

We experimentally verified that the approximation error is within 1%, comparing the result of Eq. (2) to that obtained with real simulation of (a, b, c) in every combination of $1 \le a \le 2$, $1 \le b \le 2$ and $1 \le c \le 3$. The total computational time is much reduced, although this approximation requires four thermal analyses with $(1,1,1)$, $(a,1,1)$, $(1,b,1)$ and $(1,1,c)$. For example, the CPU time for four simulations, $(1,1,1)$, $(2,1,1)$, $(1,2,1)$, and $(1,1,3)$, is 0.70 hours, while that of $(2,2,3)$ is 11.2 hours.

The result of (∞,∞,∞) could be estimated using Eq. (2), but the calculation of ∞ discretization, such as $(\infty,1,1)$, is still impossible. Therefore, considering the trade-off between accuracy and computational time, we use the result of $(2,2,3)$ approximated with Eq. (2) in this paper. For example in Fig. 7, (9, 9, 9) seems to be converged and is expected to be close to (∞,∞,∞). ΔT_{max} of (9, 9, 9) is 38.5°C and that of $(2,2,3)$ is 39.8°C by Eq. (2), and hence the difference between $(2,2,3)$ and (∞,∞,∞) is 3.5%. As mentioned in this section, the approximation error of Eq. (2) is supposed to be within 1%. Therefore, the estimation error using Eq. (2) with $(2,2,3)$ would be less than 4.5% from the value of (∞,∞,∞).

3.1.3 Update of Cu Resistivity

The Cu resistivity ρ suffers from barrier metal and scattering, especially in narrow wires [18]–[21]. ρ shown in these reports is consistent with that in the ITRS prediction, and therefore we use ρ in the ITRS roadmap. Additionally, considering the temperature dependence of Cu resistivity, the total resistivity of the copper ρ_{total} is separated into two parts: a temperature-dependent part (the phonon contribution part) ρ_{phonon}, and a temperature-independent part (the defect part) ρ_{defect} [18], [19]. If the wire temperature is T_m, $\rho_{total}(T_m)$ is expressed as Eq. (3).

$$\begin{aligned}\rho_{total}(T_m) =& \rho_{defect} + \rho_{phonon}(T_m) \\ =& \rho_{defect} + \Big\{\rho_{phonon}(T_0) + \rho_0 \cdot \mathrm{TCR}_0 \cdot (T_m - T_0)\Big\},\end{aligned} \quad (3)$$

where ρ_0 is the resistivity of the bulk material and TCR_0 is the corresponding linear temperature coefficient of resistance. T_0 is the base temperature for calculation (=20°C) and $\rho_{defect} + \rho_{phonon}(T_0)$ is ρ described in the ITRS prediction. ρ_0 and TCR_0 depend on temperature, but these variations are little in the temperature range of interest (=0 ~ 120°C). Therefore, we consider these two parameters to be constants of 2.2 [$\mu\Omega$-cm] from ITRS and 4.33×10^{-3} [K^{-1}] at 20°C.

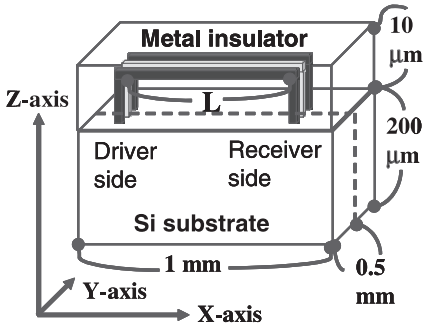

Fig. 9 A model of wire and chip used to predict self-heating in local/global signal wire.

3.2 Model of Wire and Chip

This section shows structure models for the prediction.

3.2.1 Device Parameters in Future Processes

Before explaining a model of wire and chip, we first show device parameters used for prediction in Table A· 1 (see Appendix). These values are identical to ITRS 2005 [13].

An additional parameter set that we determined is listed in Table A· 2 (also see Appendix). We select an insulating material that satisfies the value of effective dielectric constant predicted in ITRS and whose thermal conductivity is as high as possible. Both relative permittivity and thermal conductivity of the porous silica depend on its porosity. Therefore, we convert the effective dielectric constant into the porosity and then calculated the thermal conductivity [21], [22]. In following subsections, we will describe circuit models and how other additional parameters in Table A· 2 are determined.

3.2.2 Local Signal Wire Analysis

We assume an intra-block clock distribution on a high-performance chip in order to evaluate the self-heating problem in signal wires from a practical point of view. In fact, it is one of the practical worst-case scenarios, because the clock distribution involves two transitions per clock cycle. The model of wire and chip used for the evaluation are illustrated in Fig. 9. We focus on local heat diffusion immediately around the wires, and hence a 1-mm^2 chip is large enough to accurately estimate the self-heating effect.

Our purpose is to shed light on the threat of the local warming, and therefore a simple structure with three wires is selected for our analysis. Though thermal analysis under other complex configurations is straightforward, it is beyond the focus of this paper. Three intermediate wires are placed on the M5 layer in a 6-layer wiring structure including a contact layer, referring to a 90-nm industrial process technology. The center wire is driven by a 16x buffer whose output resistance is 50 Ω, and a 64x buffer (16x inverter + 64x inverter) is connected to the sink as a receiver in each

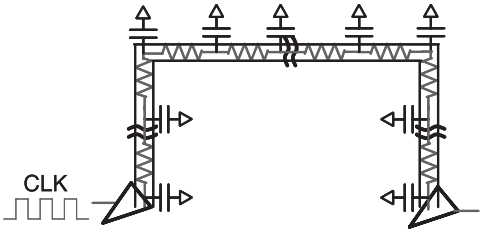

Fig. 10 Equivalent circuit used for electrical simulation with local/global signal wire.

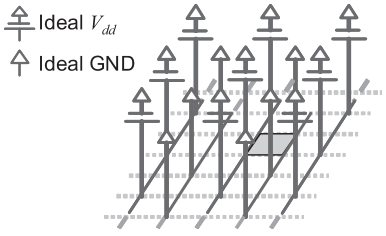

Fig. 11 A model of wire and chip used to predict self-heating in local power wire.

technology. The other two wires are configured as shielding wires for crosstalk noises, and the spacing is equal to the wire width. We assume that no heat is generated in shielding wires. Thermal coupling is analyzed in [7], but we verified that the temperature rise due to self-heating is dominant in our configuration, even if the shielding wires are active.

Figure 10 illustrates an equivalent circuit used for electrical simulation. A current flow in the wire is evaluated by circuit simulation with a predictive transistor model based on ITRS [23]. A pair of a resistor and a capacitor corresponds to one grid in the thermal simulation. Using BACPAC analytical formulae [24], wire resistances and capacitances are calculated with ITRS parameters shown in Sect. 3.2.1. The frequency of the injected clock signal is consistent with the local clock frequency in ITRS, and the input transition time is set to 10% of the clock cycle. L is the horizontal length, which does not include the vertical length from the substrate associated with wire thickness and via height. At each technology node, L is selected so that the receiver can receive a signal whose transition time is equal to one-seventh of the clock cycle, where one-seventh is an empirical parameter often used in actual designs.

Thermal circuit simulation is performed with the whole structure in Fig. 9. As realistic boundary conditions for a package, we assume that the increase in temperature of a high-performance chip is below 64°C. ITRS predicts that until 2020, the maximum affordable chip size for production and power consumption remain constant at 310 mm^2 and 198 W for high-performance chips. Hence, the power density of a chip corresponds to 0.64 (= 198/310) W/mm^2 at any process. As mentioned above, the chip size is 1-mm^2. Therefore, using our boundary conditions shown in Table A·2, the temperature increase of a chip is calculated: $0.64\times10^6/(2500+7500+1\times4) \simeq 64$ [°C], where values of 2500, 7500, and 1 correspond to the heat transfer coefficients of top, bottom, and a lateral side of the package. Recall that in our thermal analysis, we obtain power consumption with the electrical circuit simulation shown in Sect. 3.1.1, and 0.64 W/mm^2 is not used for the thermal simulation.

3.2.3 Global Signal Wire Analysis

Global wires are placed on the M11 layer of an 11-layer wiring chip, because ITRS 2005 shows that "Number of metal levels" is over 11 after 2005. For simplicity, a wire

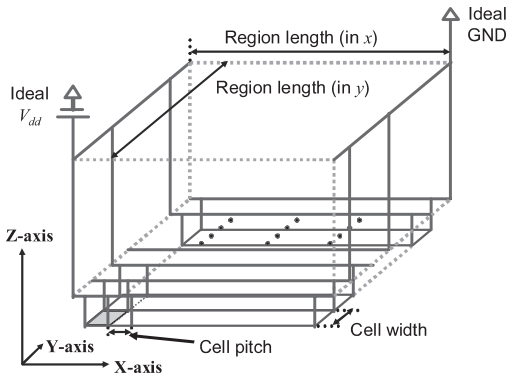

Fig. 12 Magnified view of filled region in Fig. 11.

width is assumed to be half the width of "Minimum global wiring pitch" in Table A·1. The model of wire and chip is the same as that in Fig. 9 and Fig. 10. The global clock frequency is assumed to be one-fourth of the local clock frequency, because a global clock with a higher frequency is precluded by the large interconnect delay. If the global clock were the same as the local clock, the temperature would increase roughly by a factor of four because the thermal analysis is done with the linear circuit.

To clearly illustrate the self-heating problem in global wires, a previous work [8] assumed an extreme configuration of a current. The impact of self-heating in a practical design with a buffer insertion, however, is still unclear. Therefore, we here consider the optimal repeater-inserted global wire. Buffer size and wire length L are calculated by an optimal solution of repeater insertion for minimizing propagation delay [25], [26], using ITRS parameters. This buffer insertion precludes excessively long wire configurations. Nonetheless, L of a global signal wire is much larger than L of a local signal wire.

3.2.4 Local Power Wire Analysis

In digital circuits, local power distribution network is regularly structured, and here a typical structure in Fig. 11 is analyzed. Thanks to its regular and symmetric structure, we only analyze a unit structure shown in Fig. 12 with a side boundary condition of thermal insulation. Heat transfer coefficients of top and bottom surfaces are calculated in each process technology so that the increase in temperature of the Si substrate surface from the ambient temperature becomes

up to 60°C, when the circuit operates as described in the next paragraph.

Power and ground wires on the M1 layer are configured in double back style and are connected to buffers. Ideal V_{dd} and GND are connected at the M4 layer. Power/ground wire width, cell pitch and cell width are determined by reference to a 90-nm industrial cell library. Each cell contains a driver connected to a receiver. These driver and receiver are the same as those used in Sect. 3.2.2. The input signal given to the buffers is the same as that of local signal wire analysis except the switching activity is set to 20%. Region lengths in (x, y) direction are determined in order to satisfy the condition that maximum IR drop equals 10% of V_{dd}.

4. Prediction Results

We here evaluate ΔT, which is the temperature rise from the temperature of the Si substrate surface under the center of the chip. In Sect. 4.1, we evaluate the local signal wires. Section 4.1.1 demonstrates prediction results for self-heating. Section 4.1.2 investigates which scaling parameter dominantly determines the trend of the temperature increase. The prediction results for global signal wires and local power wires are compared with those for the short intra-block wire in Sect. 4.2 and Sect. 4.3. The impact of increased temperature on reliability and performance is analyzed in Sect. 4.4. Section 4.5 examines the results.

4.1 Local Signal Wire Analysis

4.1.1 Increase in Temperature with Process Scaling

Figure 13 shows how ΔT_{max} increases with process scaling, where ΔT_{max} is the maximum temperature difference between the Si substrate and the hottest point in the wire. The predicted results indicate that self-heating in local signal wires will be significant as technology advances. ΔT_{max} increases to 40.4°C at 14 nm, while it is 0.90°C at 90 nm. It is reasonable that self-heating in local signal wires has not been considered so far, but, it could be large in the future.

ΔT distributions in 2008 (Tech. node: 59 nm) and 2020 (Tech. node: 14 nm) are depicted in Figs. 14 and 15. It can be seen that a left side of a wire, which is close to the driver, undergoes a greater increase in temperature in Fig. 14. This can be attributed to the larger current flow at the driving point, as explained in Fig. 5. ΔT in Fig. 15 seems to have a symmetrical distribution, because L is very short and the receiver input capacitance has more dominant effect on the increase in temperature than the wire capacitance. But in a closer view, a part near the driver has slightly higher temperature. These figures indicate that the generated heat diffuses through not only the via but also the interlevel metal insulator. Even for short wires, the heat diffusion characteristics through the insulator are important. We think that the increase in thermal resistance due to the smaller wire cross-sectional area also causes the increase in temperature, as shown in Sect. 2. Further analysis will be presented in the next subsection.

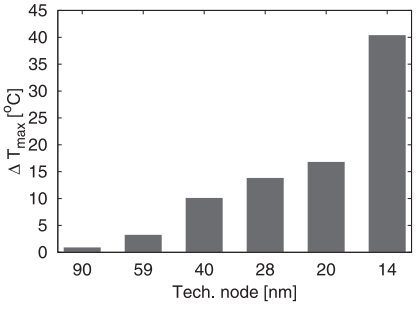

Fig. 13 Prediction of ΔT_{max} (local signal wire).

4.1.2 Attribution Analysis of Temperature Rise

This section examines which process parameter dominantly influences the self-heating effect. The evaluation is performed such that one of the factors is changed and the other factors are fixed to those in a 14-nm process. The factors to examine are the following parameters (1)–(4): (1) distance from substrate, (2) thermal conductivity of insulating material, (3) power consumption, (4) wire cross-sectional area. Figure 16 shows the result. The vertical axis denotes "$Contribution_{(i)}$" defined in Eq. (4), which means how much the parameter (i) contributes to temperature change.

$$Contribution_{(i)}\ [^\circ\mathrm{C}] = \Delta T_{max(i)} - \Delta T_{max(i)}^{(90\,\mathrm{nm})}, \tag{4}$$

where $\Delta T_{max(i)}$ denotes ΔT_{max} in the case of changing the

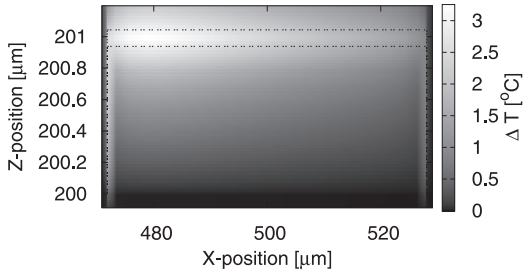

Fig. 14 Distribution of ΔT in 2008 with local signal wire (xz plane, y-coordinate is the center of the chip). Axes are indicated in Fig. 9. Region inside dotted lines corresponds to wire.

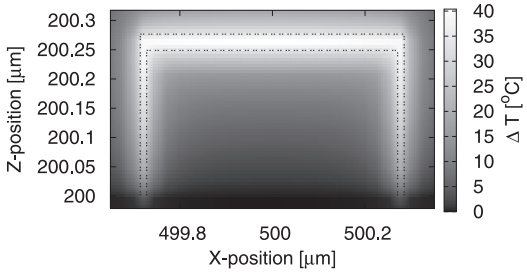

Fig. 15 Distribution of ΔT in 2020 with local signal wire (xz plane, y-coordinate is the center of the chip). Axes are indicated in Fig. 9. Region inside dotted lines corresponds to wire.

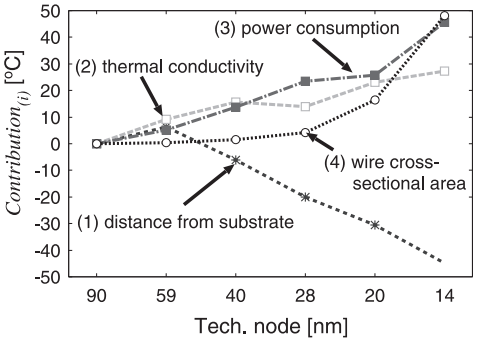

Fig. 16 Attribution analysis of ΔT_{max} (local signal wire).

parameter (i) and $\Delta T_{max(i)}^{(90\,\mathrm{nm})}$ is $\Delta T_{max(i)}$ in a 90-nm process. We can see that $\Delta T_{max(i)}$ decreases with (1), and increases with (2)–(4). Considering all the factors together, we found that the effect of the increase in temperature is larger than that of the decrease in temperature, which approximately results in the temperature rise predicted in Sect. 4.1.1. A similar experiment was also conducted for wire length L, but it is difficult to compare its *Contribution*s and we here omit it. This is the reason that the sum of all contributions in Fig. 16 differs slightly from ΔT_{max} in Fig. 13.

As Fig. 16 indicates, ΔT_{max} decreases as the distance between the wire and the substrate diminishes, and increases with lower-κ and larger power consumption. These facts are well known. However, it is notable that the size of the cross-sectional area affects self-heating as much as insulating material and power consumption below a 28-nm process. The temperature increase due to the reduction of the cross-section is rapidly growing as technology advances, which means the cross-sectional area has a large impact in the future. Although the poor thermal conductance of future insulating materials has been studied in the literature [8], [21], the influence of a smaller cross-sectional area on self-heating, which we point out here, has not been paid attention to. Smaller wire configurations degrade thermal conductivity as well as electrical conductivity.

4.2 Global Signal Wire Analysis

Here, we predict self-heating in global signal wires, and compare the results with those for local signal wires.

4.2.1 Increase in Temperature with Process Scaling

The trend in ΔT_{max} is shown in Fig. 17. Figure 18 presents ΔT distribution in 2020 (Tech. node: 14 nm). These results show that the self-heating effect in global signal wires will be very small in designs that minimize propagation delay. The increase in temperature of local signal wires will grow much more than that of global signal wires.

Repeaters are not always optimally inserted in actual designs because of limitations such as layout and power consumption. Generally speaking, the smaller the number of repeaters is, the less power is dissipated. Reference [27] reports that reducing the number of repeaters by 50% from the optimal number for power minimization incurs a 10% delay penalty. That is, when the wire length is twice as long as the optimal length, the propagation delay increases by 10%. In such designs, for example, the ΔT_{max} trend becomes "Global (2x length)" in Fig. 17. The increase in temperature becomes larger but is still smaller than that of local wires. Figure 19 shows ΔT distribution in a horizontal part of the wire with regard to those two configurations in a 14-nm process. Designs with further longer wires are less practical because of the performance degradation. Thus, our results indicate that self-heating in local wires will be a more important problem than that in global wires.

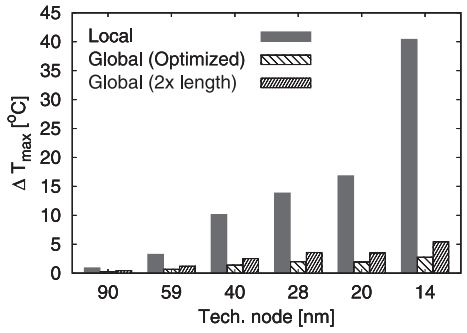

Fig. 17 Prediction of ΔT_{max} of global signal wire. Global wires are optimally buffer-inserted or are twice as long as the optimal wire. Bars labeled "Local" are the same as in Fig. 13.

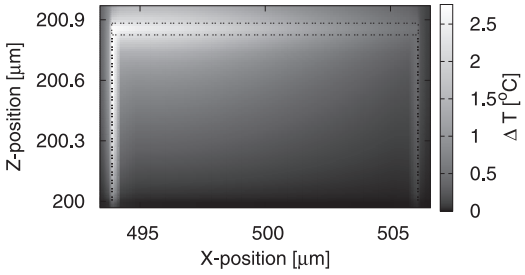

Fig. 18 Distribution of ΔT in 2020 with optimized global signal wire (xz plane, y-coordinate is the center of the chip). Axes are indicated in Fig. 9. Region inside dotted lines represents wire.

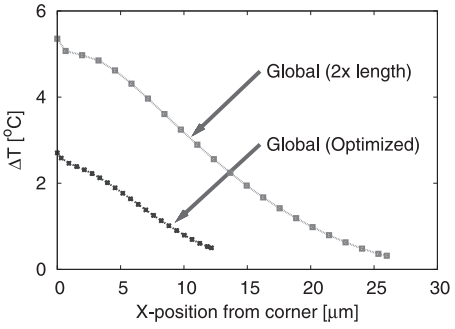

Fig. 19 Distribution of ΔT in 2020 with global signal wire (in horizontal part of wire).

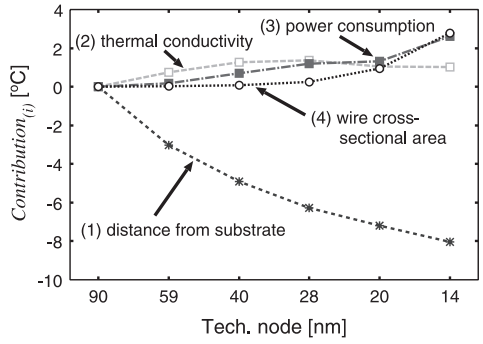

Fig. 20 Attribution analysis of ΔT_{max} (global wire).

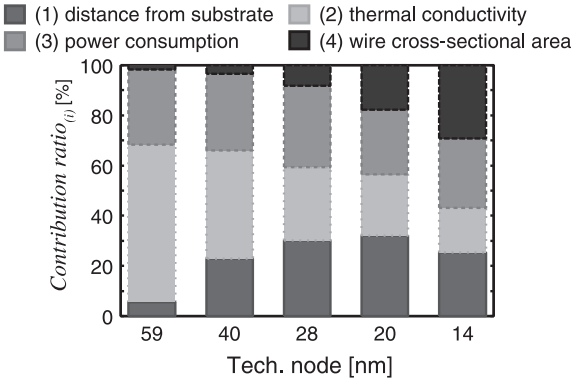

Fig. 21 *Contribution ratio*$_{(i)}$ (local signal wire).

4.2.2 Attribution Analysis of Temperature Rise

We perform the attribution analysis for the optimal repeater-inserted global wire in Sect. 4.2.1 considering technology scaling in the similar way as Sect. 4.1.2. The result is shown in Fig. 20 and its trend is almost similar to that of local wires (Fig. 16). To make a comparison easy, we define *Contribution ratio*$_{(i)}$ as Eq. (5).

$$\textit{Contribution ratio}_{(i)} = \frac{|\textit{Contribution}_{(i)}|}{\sum_i |\textit{Contribution}_{(i)}|}. \quad (5)$$

Contribution ratio$_{(i)}$ shows how much effect the parameter (i) has on the sum of temperature variations. In order to normalize the results to 100%, we here use absolute values of *Contribution*$_{(i)}$.

Figures 21 and 22 show *Contribution ratio*$_{(i)}$ trends of local and global signal wires, respectively. The result at 90 nm is omitted because every *Contribution*$_{(i)}$ equals 0 at 90 nm, as is clear from Eq. (4). We can see that (1); distance from substrate affects the self-heating dominantly in the global wiring. By contrast, (2) to (4), i.e. thermal conductivity of insulating material, power consumption, and wire cross-sectional area, have larger effects in local wires than in global wires. This results show that the distance from substrate is a key parameter that distinguishes between local and global wires in the context of self-heating. The contribution of wire cross-section area is becoming larger in global wires as well as in local wires, though the impact in local wires is larger.

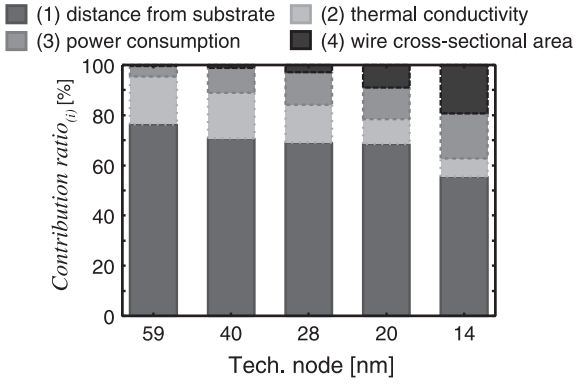

Fig. 22 *Contribution ratio*$_{(i)}$ (global signal wire).

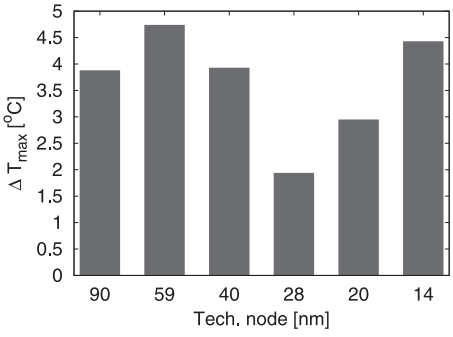

Fig. 23 Prediction of ΔT_{max} (local power wire).

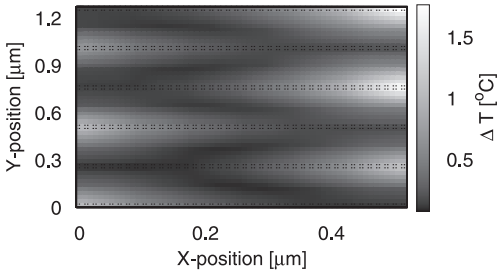

Fig. 24 Distribution of ΔT in 2020 with local power wire (xy plane, z-coordinate is the M1 layer). Axes are indicated in Fig. 12. Region inside dotted lines corresponds to wire.

4.3 Local Power Wire Analysis

The trend in ΔT_{max} of local power wires is shown in Fig. 23. Figures 24 and 25 present ΔT distributions of the 2020 device in xy and xz plane, respectively. The reason why the increases in temperature are below 5°C seems that the generated heat flows into the Si directly through a thin dielectric layer with a small thermal resistance, and thus the temperature of the whole chip rises. As long as self-heating is evaluated with ΔT, which is the temperature increase from the center of a chip, it is quite small.

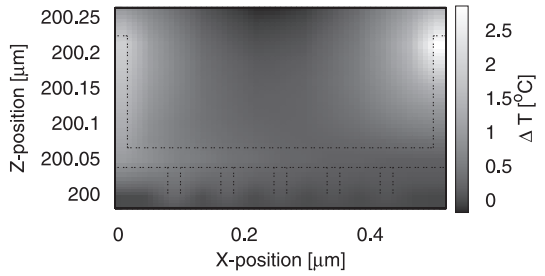

Fig. 25 Distribution of ΔT in 2020 with local power wire (xz plane, y-coordinate is the center of the region). Axes are indicated in Fig. 12. Region inside dotted lines corresponds to wire.

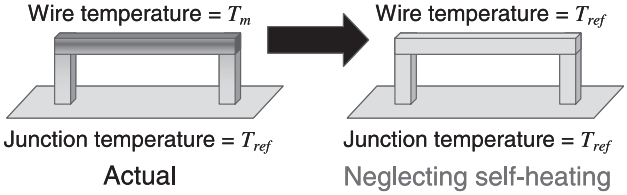

Fig. 26 Neglecting self-heating in a wire.

4.4 Impact of Self-Heating

We quantitatively evaluate the unexpected degradations of reliability and performance for the case that self-heating of local wires is neglected and the metal wire temperature T_m is assumed to be the same as the reference chip temperature at the Si junction T_{ref}. Figure 26 illustrates this assumption. For simplicity, we here assume $T_m = T_{ref} + \Delta T_{avg}$ throughout a wire. ΔT_{avg} is calculated in Eq. (6),

$$\Delta T_{avg} = \frac{\sum_{i \in \text{center wire}} \Delta T_i \times Vol_i}{\sum_{i \in \text{center wire}} Vol_i}, \tag{6}$$

where ΔT_i and Vol_i denote ΔT and the volume of mesh i in the center wire, respectively. In this section, we evaluate a 14-nm case in which the temperature increase was largest in our experiment.

4.4.1 Electromigration Reliability

Electromigration (EM) is one of the major failure mechanisms in interconnects under unidirectional current stress [1]. EM lifetime reliability of a metal wire is usually modeled by Black's equation,

$$\text{MTTF} = A^* j^{-n} \exp\left(\frac{Q}{k_B T_m}\right), \tag{7}$$

where MTTF is the mean time to failure (typically for 0.1% cumulative failure), A^* is a constant that is dependent on the geometry and microstructure of the interconnect, j is the DC or average current density, and n is 2 under normal use conditions. Activation energy Q is dominated by surface transport in narrow ($< 1\,\mu$m) copper lines [28] and is approximately 0.5 eV (=8.01089×10^{-20} [J]) [29]. k_B is Boltzmann's constant (=1.38065×10^{-23} [JK^{-1}]), and T_m is the metal temperature. We evaluate how optimistically MTTF can be estimated when self-heating of local wires is neglected, i.e. wire temperature T_m [K] is set to silicon junction temperature T_{ref} [K]. In this case, the estimated MTTF is $\frac{\text{MTTF}_{T_{ref}}}{\text{MTTF}_{T_m}}$ times larger than the actual MTTF with T_m,

$$\frac{\text{MTTF}_{T_{ref}}}{\text{MTTF}_{T_m}} = \frac{A^* j^{-2} \exp(\frac{Q}{T_{ref} k_B})}{A^* j^{-2} \exp(\frac{Q}{T_m k_B})} = \exp\left(\frac{Q}{K_B} \cdot \frac{T_m - T_{ref}}{T_m T_{ref}}\right). \tag{8}$$

Using the result of the 14-nm process in Sect. 4.1.1, ΔT_{avg} = 32.4°C. The actual MTTF is 3.08 (T_{ref} = 120°C) to 9.51 (T_{ref} = 0°C) times shorter than the MTTF estimated with T_{ref}. However, for signal wires and clock distribution network, the thermal impact on EM is of less concern, because the current is bidirectional and the MTTF is inversely proportional to the clock frequency [30]. We next consider the worst case that a large DC current flows throughout a wire. The DC current density assumes to be the maximum current density J_{max} predicted by ITRS, which is 27.4 [MA/cm^2] in 2020. Using 2D simulation similar to that in Sect. 2, ΔT_{avg} at 14 nm results in 14.4°C. By contrast, for local power wires, ΔT_{avg} is calculated using the result in Sect. 4.3 and is approximately 0.5°C.

The results of Eq. (8) with local signal and power wire at 14 nm are shown in Table 2. The actual MTTF is at worst approximately 3 times shorter than the MTTF estimated with T_{ref}. When an EM margin given to the wire is small, the overestimation of MTTF may cause an unexpected reliability failure.

Table 2 Estimated MTTF mismatch $\left(\frac{\text{MTTF}_{T_{ref}}}{\text{MTTF}_{T_m}}\right)$ at 14 nm.

Wire	T_{ref} = 0°C	T_{ref} = 120°C
Local signal	2.90	1.68
Local power	1.04	1.02

4.4.2 Performance Degradation

As is well known, interconnect delay is expressed by Eq. (9) [31],

$$\begin{aligned}\text{Delay}_{50\%}(T_{tr}, T_{wire}) &= R_{tr}(T_{tr}) \cdot (0.693 C_{int} + 0.693 C_L) \\ &+ R_{int}(T_{wire}) \cdot (0.377 C_{int} + 0.693 C_L),\end{aligned} \tag{9}$$

where T_{tr} is driver temperature, T_{wire} is wire temperature, R_{tr} is driver output resistance, R_{int} is wire resistance, C_{int} is wire capacitance, and C_L is fanout load. The capacitances are independent of temperature. As mentioned in Sect. 3.1.3, the resistivity of the copper is expressed as Eq. (3). Therefore, delay estimation error $\Delta\text{Delay}_{50\%}$ due to self-heating is calculated in Eq. (10).

$$\begin{aligned}&\Delta\text{Delay}_{50\%}(T_{ref}, T_m) \\ &= \text{Delay}_{50\%}(T_{ref}, T_m) - \text{Delay}_{50\%}(T_{ref}, T_{ref})\end{aligned}$$

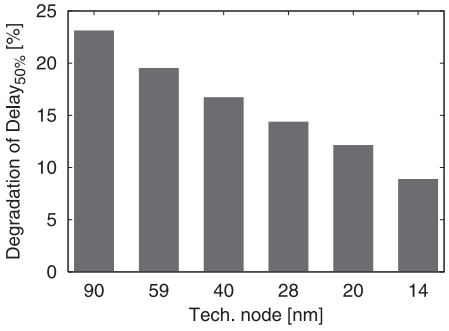

Fig. 27 Diminishing dependence of temperature on $\text{Delay}_{50\%}$ (T_m = 120°C, T_{ref} = 27°C). The value on vertical axis is calculated by means of $\Delta\text{Delay}_{50\%}(T_{ref}, T_m)/\text{Delay}_{50\%}(T_{ref}, T_{ref})$.

$$= \left\{R_{int}(T_m) - R_{int}(T_{ref})\right\} \cdot (0.377C_{int} + 0.693C_L). \tag{10}$$

Using ΔT_{avg} of 32.4°C in local signal wire and T_{ref} of 27°C, $\text{Delay}_{50\%}(T_{ref}, T_{ref})$ is 303 fs and $\Delta\text{Delay}_{50\%}(T_{ref}, T_m)$ is 9.39 fs at a 14-nm process. It indicates that degradation in performance is 3.10%.

4.5 Discussion

Above results show that self-heating of local signal wire is larger than those of global signal wire and local power wire. If no breakthroughs will come up in the design, what we can do for local signal wires is only to keep a wire cross-sectional area large. From the attribution analysis in Sect. 4.1.2, applying a 90-nm wire width and height to all technologies enables us to keep ΔT_{max} below 10°C.

On the other hand, in contrast to our expectations, the performance of local wires can be reasonably estimated with the temperature of substrate. The reason why degradation in performance is not significant is that the dependence of temperature on wire resistance is relatively small because of the size effect explained in Eq. (3). When T_m=120°C and T_{ref}=27°C, for example, ρ increases by 28.2% (from 3.14 to 4.02 $\mu\Omega$-cm) at a 90-nm process, and by 10.7% (from 8.26 to 9.14 $\mu\Omega$-cm) at a 14-nm process. The resistivity increase due to size effect is not dependent on temperature, and hence the resistivity will be less temperature-dependent. Thus, the delay degradation is decreasing with process scaling. Figure 27 shows performance degradation for the case that T_m=120°C and T_{ref}=27°C. It can be seen that the degradation in performance becomes less sensitive to the increase in temperature with process scaling.

The predictive results in this work come to the following remarks: (1) The temperature difference between the wire and the substrate will become large in advanced technologies, especially at a 14-nm process, when a local wire propagates a signal with high operating frequency and switching activity, such as clock. (2) Reducing the area of wire cross-section exacerbates self-heating as well as using low-κ material. (3) The wire delay will become less sensitive to the change in temperature. However, when a wire cross-section is small and large current flows, the wire reliability may degrade unexpectedly due to self-heating.

Further consideration of the following points would be beneficial for practical designs: (1) Transient thermal analysis could provide design implications, although steady-state analysis is sufficient in this paper because the evaluated circuits have constant activity (100%). For example, when a circuit block wakes up from suspend/sleep mode, large current flows, and it may cause temporally- and spatially-local elevation of temperature. (2) Development of a comprehensive design guideline considering both of substrate temperature and wire temperature is helpful especially for reliability guardbanding.

5. Conclusion

We predicted that the self-heating effect in local wires will become apparent with process scaling. Our results show that the temperature in local signal wires will grow more drastically than that in the optimal repeater-inserted global wires as technology advances. The cross-sectional area of a wire affects self-heating as significantly as the insulating material and power consumption in advanced technologies below 28 nm. When local wires do not have a large EM margin, underestimating their temperature by neglecting the self-heating effect may cause an unexpected reliability failure. By contrast, self-heating in local wires has a limited impact on their performance.

Acknowledgments

This work was supported in part by NEDO.

References

[1] J.R. Black, "Electromigration — A brief survey and some recent results," IEEE Trans. Electron Devices, vol.ED-16, no.4, pp.338–347, April 1969.
[2] K. Banerjee, M. Pedram, and A.H. Ajami, "Analysis and optimization of thermal issues in high-performance VLSI," Proc. ISPD, pp.230–237, 2001.
[3] K. Skadron, M.R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-aware computer systems: Opportunities and challenges," IEEE Micro, vol.23, no.6, pp.52–61, Nov.–Dec. 2003.
[4] Y.K. Cheng, C.H. Tsai, C.C. Teng, and S.M. Kang, Electrothermal Analysis of VLSI Systems, Kluwer Academic Publishers, 2000.
[5] A.H. Ajami, K. Banerjee, and M. Pedram, "Modeling and analysis of nonuniform substrate temperature effects on global ULSI interconnects," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol.24, no.6, pp.849–861, June 2005.
[6] S. Im, K. Banerjee, and K.E. Goodson, "Modeling and analysis of via hot spots and implications for ULSI interconnect reliability," Proc. IRPS, pp.336–345, April 2002.
[7] T.Y. Chiang, K. Banerjee, and K.C. Saraswat, "Compact modeling and SPICE-based simulation for electrothermal analysis of multilevel ULSI interconnects," Proc. ICCAD, pp.165–172, 2001.
[8] K. Banerjee and A. Mehrotra, "Global (interconnect) warming," IEEE Circuits Devices Mag., vol.17, no.5, pp.16–32, Sept. 2001.
[9] T.Y. Chiang, K. Banerjee, and K.C. Saraswat, "Effect of via separation and low-k dielectric materials on the thermal characteristics of

Cu interconnects," Proc. IEDM, pp.261–264, Dec. 2000.
[10] K. Banerjee, A. Mehrotra, A.S. Vincentelli, and C. Hu, "On thermal effects in deep sub-micron VLSI interconnects," Proc. DAC, pp.885–891, June 1999.
[11] S. Rzepka, K. Banerjee, E. Meusel, and C. Hu, "Characterization of self-heating in advanced VLSI interconnectlines based on thermal finite element simulation," IEEE Trans. Compon. Packag. Manuf. Technol., vol.21, no.3, pp.406–411, Sept. 1998.
[12] K. Shinkai, M. Hashimoto, and T. Onoye, "Future prediction of self-heating in short intra-block wires," Proc. ISQED, pp.660–665, March 2007.
[13] http://www.itrs.net/
[14] C.H. Tsai and S.M. Kang, "Cell-level placement for improving substrate thermal distribution," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol.19, no.2, pp.253–266, Feb. 2000.
[15] Mathematical Systems, Inc., "LiCRSIM Users Manual."
[16] C. Mizuta, J. Iwai, K. Machida, T. Kage, and H. Masuda, "Large-scale linear circuit simulation with an inversed inductance matrix," Proc. ASP-DAC, pp.511–516, Jan. 2004.
[17] ANSYS, Inc., "ANSYS Users Manual."
[18] W. Steinhögl, G. Schindler, G. Steinlesberger, M. Traving, and M. Engelhardt, "Comprehensive study of the resistivity of copper wires with lateral dimensions of 100 nm and smaller," J. Appl. Phys., vol.97, no.2, pp.023706-1–023706-7, 2005.
[19] W. Steinhögl, G. Schindler, G. Steinlesberger, and M. Engelhardt, "Scaling laws for the resistivity increase of sub-100 nm interconnects," Proc. SISPAD, pp.27–30, Sept. 2003.
[20] W. Steinhögl, G. Schindler, G. Steinlesberger, and M. Engelhardt, "Size-dependent resistivity of metallic wires in the mesoscopic range," Phys. Rev., vol.66, no.7, pp.075414-1–075414-4, Aug. 2002.
[21] S. Im, N. Srivastava, K. Banerjee, and K.E. Goodson, "Scaling analysis of multilevel interconnect temperatures for high-performance ICs," IEEE Trans. Electron Devices, vol.52, no.12, pp.2710–2719, Dec. 2005.
[22] B.Y. Tsui, C.C. Yang, and K.L. Fang, "Anisotropic thermal conductivity of nanoporous silica film," IEEE Trans. Electron Devices, pp.20–27, 2004.
[23] S. Uemura, A. Tsuchiya, and H. Onodera, "A predictive transistor model based on ITRS," http://www.tamaru.kuee.kyoto-u.ac.jp/˜tsuchiya/tr-model.html.en
[24] http://www.eecs.umich.edu/˜dennis/bacpac/models/delay.html
[25] H.B. Bakoglu, Circuits, Interconnections, and Packaging for VLSI, Addison-Wesley Publishing Company, 1990.
[26] R.H.J. M Otten and R.K. Brayton, "Planning for performance," Proc. DAC, pp.122–127, 1998.
[27] J.A. Davis and J.D. Meindl, Interconnect Technology and Design for Gigascale Integration, Springer (Kluwer Academic Publishers), 2003.
[28] C.K. Hu, R. Rosenberg, and K.Y. Lee, "Electromigration path in Cu thin-film lines," Appl. Phys. Lett., vol.74, no.20, pp.2945–2947, May 1999.
[29] B.H. Jo and R.W. Vook, "In-situ ultra-high vacuum studies of electromigration in copper films," Thin Solid Films, vol.262, no.1–2, pp.129–134, 1995.
[30] J. Tao, N.W. Cheung, and C. Hu, "Modeling electromigration lifetime under bidirectional currentstress," IEEE Electron Device Lett., vol.16, no.11, pp.476–478, Nov. 1995.
[31] T. Sakurai, "Closed-form expressions for interconnection delay, coupling, and crosstalk in VLSIs," IEEE Trans. Electron Devices, vol.40, no.1, pp.118–124, Jan. 1993.

Appendix: Device Parameters

Tables A·1 and A·2 show device parameters used in this work. "Analysis" column indicates for which analysis parameters are used. LS, GS, and LP stand for **L**ocal **S**ignal, **G**lobal **S**ignal, and **L**ocal **P**ower.

Table A·1 Parameter set from ITRS.

Analysis	Parameter	Year (Tech. node [nm])					
		2005 (90)	2008 (59)	2011 (40)	2014 (28)	2017 (20)	2020 (14)
LS/GS/LP	Aspect ratio of Metal 1	1.7	1.8	1.9	1.9	2.0	2.0
	Aspect ratio of intermediate (wire)	1.7	1.8	1.9	1.9	2.0	2.0
	Aspect ratio of intermediate (via)	1.5	1.6	1.6	1.7	1.8	1.8
	Poly-Si or metal gate electrode thickness [nm]	64	46	32	22	16	12
	Tungsten contact plug height [nm]	135	94.4	64	47.6	36	25.2
	V_{dd} [V]	1.1	1.0	1.0	0.9	0.7	0.7
	Local clk [GHz]	5.204	10.972	17.658	28.356	45.535	73.122
	Conductor effective resistivity (Cu Metal 1 wiring) [$\mu\Omega$-cm]	3.15	3.67	4.3	5.2	6.33	8.19
	Conductor effective resistivity (Cu intermediate wiring) [$\mu\Omega$-cm]	3.07	3.65	4.3	5.2	6.33	8.19
	Effective dielectric constant	3.1	2.7	2.5	2.4	1.9	1.6
GS	Minimum global wiring pitch [nm]	300	177	120	84	60	42
	Aspect ratio of global (wire)	2.2	2.3	2.4	2.5	2.6	2.8
	Aspect ratio of global (via)	2	2.1	2.2	2.3	2.4	2.5
	Conductor effective resistivity (minimum pitch Cu global wiring) [$\mu\Omega$-cm]	2.53	2.87	3.22	3.73	4.39	5.38
	$I_{d,sat}$ (NMOS) [μA/μm]	1020	1570	2490	2290	2533	2981
	$I_{d,sat}$ (PMOS) [μA/μm]	408	628	996	916	1013.2	1192.4
	$C_{g,total}$ [fF/μm]	0.813	0.847	0.859	0.542	0.487	0.362

Table A· 2 Additional parameter set (determined in this work).

Analysis	Parameter		Year (Tech. node [nm]) 2005 (90)	2008 (59)	2011 (40)	2014 (28)	2017 (20)	2020 (14)
LS/GS/LP	Interlevel metal insulator (Thermal conductivity [W/m·K])		FSG 0.89	SiOC 0.39	Porous silica 0.231	0.207	0.162	0.115
LS/GS	Heat transfer coefficient [W/m^{2}K]	Lateral sides	1 (insulated)					
		Top	2500					
		Bottom	7500					
LS	L (wire length in horizontal part) [μm]		160	56.0	21.7	10.6	4.85	0.54
	Distance between Metal-5 layer and substrate [μm]		1.351	0.943	0.640	0.473	0.356	0.250
GS	L (wire length in horizontal part) [μm]		308.38	143.93	76.23	39.11	22.64	12.06
	Distance between Metal-11 layer and substrate [μm]		4.612	2.993	2.084	1.530	1.144	0.829
	Buffer size		7.19x	4.43x	3.29x	3.87x	2.91x	2.14x
LP	Heat transfer coefficient [W/m^{2}K]	Top	2443	5021	10731	11345	20559	68443
		Bottom	7330	15063	32194	34036	61678	205328
	Power/ground wire width [μm]		0.110	0.076	0.052	0.038	0.029	0.020
	Cell pitch (in x direction) [μm]		0.430	0.278	0.188	0.130	0.091	0.064
	Cell width (in y direction) [μm]		1.510	0.986	0.668	0.466	0.331	0.232
	Region pitch (in x direction) [μm]		11.45	7.51	3.89	1.89	1.35	0.52
	Region width (in y direction) [μm]		14.69	7.51	3.65	2.56	1.83	1.28

Ken-ichi Shinkai received the B.E. and M.E. degrees in Information Systems Engineering from Osaka University, Osaka, Japan, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree in the Department of Information Systems Engineering at Osaka University. His major interest is variation-aware timing and reliability analysis. He is a student member of IEEE.

Masanori Hashimoto received the B.E., M.E. and Ph.D. degrees in Communications and Computer Engineering from Kyoto University, Kyoto, Japan, in 1997, 1999, and 2001, respectively. Since 2004, he has been an Associate Professor in Department of Information Systems Engineering, Graduate School of Information Science and Technology, Osaka University. His research interest includes computer-aided-design for digital integrated circuits, and high-speed circuit design. Dr. Hashimoto served on the technical program committees for international conferences including DAC, ICCAD, ASP-DAC, ICCD and ISQED. He is a member of IEEE and IPSJ.

Takao Onoye received the B.E. and M.E. degrees in Electronic Engineering, and Dr.Eng. degree in Information Systems Engineering all from Osaka University, Japan, in 1991, 1993, and 1997, respectively. He is currently a professor in the Department of Information Systems Engineering, Osaka University. His research interests include media-centric low-power architecture and its SoC implementation. He is a member of IEEE, IPSJ, and ITE-J.